

Crystal Structure of the WUSCHEL Homeodomain

-

Description and Comparative Structural Analysis of a Novel Homeodomain Subtype

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln



vorgelegt von

Claudia Koppermann

aus München, Deutschland

Köln 2018

Berichtersteller:
(Gutachter)

Prof. Dr. Wolfgang Werr

Prof. Dr. Ulrich Baumann

Tag der mündlichen Prüfung:

17. November 2017

Zusammenfassung

Die Homöodomäne ist eine altertümliche DNA-Bindedomäne, die aller Wahrscheinlichkeit nach bereits in einem gemeinsamen eukaryotischen Urahn von Tieren, Pilzen und Pflanzen vorhanden war. Trotz dieser beachtlichen evolutionären Zeiträume sind grundlegende strukturelle Merkmale und der DNA-Bindemechanismus der Homöodomäne nahezu unverändert geblieben. Dies legen bisher veröffentlichte Strukturen von Homöodomänen aus der Hefe sowie aus verschiedenen Tierarten nahe. Die Grundstruktur der Homöodomänen umfasst demnach drei α -Helizes, wobei die ersten beiden Helizes antiparallel zueinander verlaufen, zu denen die dritte, die C-terminale Helix ungefähr im rechten Winkel angeordnet ist. Typische Homöodomänen bestehen aus 60 Aminosäuren. Weicht die Anzahl der Aminosäuren aus denen die Homöodomäne aufgebaut ist von 60 ab, wird diese als „atypisch“ bezeichnet. Dabei scheinen zusätzliche Aminosäuren als eine Art extra Schleife in den verbindenden Regionen zwischen den Helizes ausgelagert zu werden, so dass deren relative Orientierung zueinander, und somit die Grundstruktur der Homöodomäne, erhalten bleibt. Jedoch zählen fast alle bisher bestimmten Strukturen zu den typischen oder einer speziellen Unterklasse der atypischen Homöodomänen. Die Diversität an bekannten Homöodomänenstrukturen ist somit äußerst begrenzt.

Die Homöodomänen Familie der WOX Proteine (WUSCHEL-related homeobox) umfaßt den Namensgeber WUSCHEL, sowie dessen nahe Verwandte, deren kodierende DNA Sequenz eine Homöobox aufweist, die der im *WUSCHEL* Gen vorhandenen ähnlich ist. Die Homöodomäne, das Genprodukt der Homöobox, besteht im Falle der WOX Familie aus 65 oder 66 Aminosäuren und zählt folglich zu der Gruppe der atypischen Homöodomänen. Vertreter der WOX Familie sind als Transkriptionsfaktoren an den verschiedensten entwicklungsbiologischen Prozessen der Pflanze beteiligt. Außerdem legen phylogenetische Untersuchungen den Schluß nahe, dass die gesamte, heute bekannte WOX Familie aus einem einzelnen Gen entstanden ist, dessen Ursprung in einzelligen grünen Algen liegt. Folglich waren Vertreter der WOX Familie von den frühesten Anfängen an ständige Begleiter der Landpflanzen und haben wahrscheinlich deren Evolution während der Kolonisation der Landmasse und somit der Erschließung neues Lebensraumes maßgeblich mitgestaltet. Andere Untersuchungen weisen auf die Entstehung teilweise abweichender Funktionsweisen innerhalb der WOX Familie hin. Manche dieser Unterschiede scheinen mit evolutionären Veränderungen der WOX Homöodomäne in Verbindung zu stehen.

Aus all diesen Gründen stellt die WOX Familie ein interessantes Modellsystem dar, anhand dessen sich die Evolution und die molekularen Grundlagen neu erworbener Funktionsweisen in

Homöodomänenproteinen erforschen lassen. Vergleiche der Aminosäuresequenzen von Homöodomänen der WOX Familie deuten an, dass wahrscheinlich einige substantielle Unterschiede zu den bisher publizierten Homöodomänenstrukturen bestehen. Diese betreffen unter anderem die Integration von zusätzlichen Aminosäuren, die an zwei Stellen der WOX Homöodomäne erfolgte, zu denen keine vergleichbare Strukturbeschreibung existiert. Des Weiteren fallen Veränderungen der Aminosäurereste an bestimmten Positionen der WOX Homöodomäne auf. Deren Pendant sind in tierschen Homöodomänen oder solchen aus der Hefe an der DNA Erkennung beteiligt und dort dementsprechend stark konserviert. Da bisher noch keine Struktur einer pflanzlichen Homöodomäne publiziert wurde, wirft die erhöhte Variabilität hinsichtlich Länge und Komposition der Aminosäuresequenz in WOX Homöodomänen die Frage auf, inwiefern Ähnlichkeiten zu anderen Homöodomänen bestehen oder ob im Laufe der Evolution vielleicht teilweise abweichende Wege in Pflanzen und Tieren beschritten wurden.

Um die Basis für weitergehende Untersuchungen zur Evolution der pflanzlicher und insbesondere der WOX Homöodomäne zu legen, wurde im Rahmen der vorliegenden Doktorarbeit die Homöodomäne (HD) des Stammzellfaktors WUSCHEL (WUS) im Modellorganismus *A. thaliana* kristallisiert. Die Struktur von WUS HD konnte mit einer Auflösung von bis zu 1.85 Å bestimmt werden und deren strukturelle Analyse im Vergleich mit bereits publizierten Homöodomänenstrukturen zeigt, dass WUS HD diverse bisher nicht beschriebene Charakteristika aufweist und somit einer neuen strukturellen Untergruppe von Homöodomänen angehört. Außerdem ermöglichte die Struktur von WUS HD die Identifizierung von bestimmten Aminosäureresten, deren Veränderung im Laufe der Evolution zu unterschiedlichen Funktionsweisen in der Familie der WOX Homöodomänen geführt haben könnten.

Abstract

The homeodomain is an ancient DNA binding domain, which is likely to have been present in a common eukaryotic ancestor of the three major lineages leading to animals, fungi and plants. Despite these considerable evolutionary distances the structures of several animal and yeast homeodomains display a remarkable degree of conservation of the general overall built, as well as regarding their manner of DNA recognition. The homeodomain fold is composed of three α -helices assembled in a characteristic way. The most C-terminal helix is positioned roughly perpendicular to the other two, which run antiparallel. This general homeodomain structure was first described in case of the so called typical homeodomains, which consist of 60 amino acid residues and whose α -helices are connected by a loop and a turn, respectively. Homeodomains composed of a number of amino acids divergent from the canonical 60 were termed atypical homeodomains. Indicated by published structures of atypical homeodomains, additional amino acids seem to be simply looped out from the regions connecting the three helices, hence preserving their relative orientation to each other and the general homeodomain fold. However, nearly all of the available structures of atypical homeodomains describe TALE (three amino acid loop extension) homeodomains, whose three extra amino acid residues adopt a characteristic and highly specific conformation. Thus the diversity in published homeodomain structures is rather low.

The *WOX* (WUSCHEL-related homeobox) homeodomain family comprises WUSCHEL and several close relatives, which are characterised by a homeodomain similar to that of WUSCHEL with respect to the amino acid sequence. Homeodomains of the *WOX* family are composed of 65 or 66 amino acids and hence belong to the class of atypical homeodomains. Members of this transcription factor family were shown to be involved in various key processes during plant development, e.g. establishing the apical-basal axis in the plant embryo, maintaining a stem cell population in the root tip and the shoot apex or during the development of the plant vasculature. Moreover all *WOX* genes share a monophyletic origin in unicellular green algae. As such members of this transcription factor family have accompanied the evolution of terrestrial plants from the very beginnings throughout the gradual colonization of the land. The results of complementation assays in plants with loss of function mutations in different *WOX* genes indicate a functional divergence between individual members of the *WOX* family, which partially may relate to evolutionary changes of the *WOX* homeodomain. Thus the *WOX* family may constitute a model system to study the molecular basis of evolutionary changes and the

acquisition of novel functions in homeodomain proteins. However, amino acid sequence alignments indicate several differences between homeodomains of the WOX family and the structures published up to now. Most noticeable is the integration of additional amino acids at two different sites of the WOX homeodomain, none of which is resembled by available structures of other atypical homeodomains. Furthermore, individual positions of the homeodomain, which were shown to be essential for DNA recognition and thus are well conserved in animal homeodomains, differ in the WOX family. The variability of chain length and amino acid composition present in WOX homeodomains also raises more fundamental questions about putative differences and similarities between the evolutionary trajectories of plant and animal homeodomains, in particular since no structure of a plant homeodomain has been determined to date.

In order to close the gap in diversity of available homeodomain structures and to facilitate a more direct approach in understanding the evolution of the WOX homeodomain, the homeodomain (HD) of the plant stem cell factor WUSCHEL (WUS) in *A. thaliana* was crystallized and the X-ray structure was obtained with a resolution up to 1.85 Å. The structural comparison of WUS HD and other homeodomains revealed several distinctive characteristics, which have not been described so far and thus identify WUS HD as a novel structural subtype. Furthermore a model of WUS HD bound to DNA facilitated the identification of strong candidate amino acid residues, which are likely to result in functional differences of WOX homeodomains.

Table of content

Zusammenfassung	3
Abstract	5
Abbreviations	9
1 Introduction	10
1.1 The homeodomain	10
1.1.1 Short survey of the homeodomain and classification by structural characteristics	10
1.1.2 General concept of homeodomain DNA recognition	12
1.1.3 Functional specificity of homeodomain monomers	15
1.1.3.1 The sequence specificity is largely defined by the amino acid composition of residues with direct contact to nucleotide bases	15
1.1.3.2 Sequence recognition of the N-terminal arm	20
1.1.3.3 Influence of backbone contacts outside the N-terminal arm on DNA recognition	23
1.2 The WUSCHEL-related homeobox (WOX) family	26
1.2.1 Short introduction on the evolution of terrestrial plants	26
1.2.2 Discovery of a novel homeodomain family, classification and evolution of its members	27
1.2.3 Functional divergence of WOX proteins	29
2 Materials and methods	32
2.1 Chemicals	32
2.2 Vectors	32
2.3 Kits	33
2.4 Oligonucleotides	33
2.5 Bacteria strains	34
2.6 Media for bacterial cell cultures	34
2.7 Buffers	35
2.8 Amino acid frequency maps	37
2.9 Gelelectrophoresis	38
2.9.1 Agarose gel electrophoresis	38
2.9.2 Polyacrylamide gel electrophoresis (PAGE)	38
2.10 Cloning	38
2.11 Transformation	39
2.12 Test expressions	40
2.13 Large scale protein expression	40
2.14 Protein purification	40
2.14.1 Cell lysis and clarification of lysate	41
2.14.2 Affinity purification via Int/CBD tag	41
2.14.3 Affinity purification via 6xHis tag	41
2.14.4 Size exclusion chromatography	42
2.15 Protein crystallisation	42
2.16 Optimization of WUSCHEL homeodomain needle cluster	43
2.17 Data collection and determination of the WUSCHEL homeodomain structure	43
2.18 Homology modelling of other WOX homeodomains	44
2.19 Structural analysis	45
3 Results	46
3.1 Description of the crystal structure of the WUSCHEL homeodomain	46
3.1.1 Overall structure	46

3.1.2	Non-crystallographic symmetry	47
3.1.3	Hydrogen bonds and electrostatic interactions	49
3.1.4	The hydrophobic core	50
3.1.5	Stabilization of the N-terminal arm and the loop in WUS HD	52
3.2	Comparative analysis of WUS HD and other homeodomain structures	54
3.2.1	Identification of amino acid residues of WUS HD without structural equivalent in typical and atypical homeodomains	54
3.2.2	Comparison of core amino acid residues	58
3.2.3	Novel structural characteristics of WUS HD	62
3.2.3.1	Stabilization of the N-terminal arm distinguishes WUS HD from other published homeodomain structures	62
3.2.3.2	The characteristic conformation of DR I results from the integration of additional amino acids and a distinctive hydrophobic core amino acid residue	64
3.2.3.3	Simultaneous extension of helix II and the adjacent turn preserve the HTH-motif in WUS HD	66
3.3	WUS HD represents a novel homeodomain subtype	67
4	Discussion	69
4.1	The <i>wus-7</i> allele and putative structural basis for impaired function of <i>wus</i> mutant protein	69
4.2	Possible effects of novel characteristics identified in WUS HD on DNA recognition	70
4.2.1	Generation of models illustrating the potential DNA recognition of WUS HD	70
4.2.2	Residues of the putative DNA interaction surface of WUS HD	73
4.2.3	Uncommon amino acids of helix III may affect sequence recognition of WUS HD	75
4.2.4	The differences in linking the N-terminal arm with the helix bundle apparent in WUS HD may affect its DNA recognition	76
4.2.5	The unique conformation of DR I of WUS HD may result in alterations of DNA backbone contacts	78
4.3	Implications on the evolution of the homeodomain structure and function in the WOX family inferred from the WUS HD structure and conservation of key amino acid residues	80
5	References	85
6	Supplementary information	91
	Danksagung	92
	Erklärung	93

Abbreviations

amino acids	are abbreviated utilizing the 3-letter code, only exception is the WFXN motif of helix III, where the 1-letter code is used
nucleotides	are abbreviated utilizing the 1-letter code
ABD-B	ABDOMINAL-B
AL	ARISTALESS
ANTP	ANTENNAPEdia
ASA	accessible surface area
BCD	BICOID
CBD	chitin binding domain
chA/B	chainA/B
CLL	CLAWLESS
EN	ENGRAILED
EXD	EXTRADENTICAL
FM	floral meristem
FTZ	FUSHI TARAZU
HD	homeodomain
HNF1	HEPATOCYTE NUCLEAR FACTOR
HOX	HOMEODOMAIN
HTH	helix-turn-helix
ITC	isothermal titration calorimetry
LSQ	least square fit
MAT	MATING TYPE
MSH	MUSCLE SEGMENT HOMEODOMAIN
MxeInt	mini-intein from the modified <i>Mycobacterium xenopi</i> <i>gyrA</i> gene
NCS	non-crystallographic symmetry
NMR	Nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PBX1	Pre-B-cell leukemia transcription factor 1
PCR	polymerase chain reaction
PDB ID	accession number (ID) of a structure deposited in the Protein Data Bank (PDB)
PITX2	PITUITARY HOMEODOMAIN PROTEIN 2
QC	quiescent centre
RAM	root apical meristem
RMSD	root mean square deviation
SAM	shoot apical meristem
SceInt	intein from the <i>Saccharomyces cerevisiae</i> VMA1 gene
SCR	SEX COMP REDUCED
SEC	size exclusion chromatography
SSM	secondary structure matching
TALE	three amino acid loop extension
TPL/WISP1	TOPLESS/WUSCHEL INTERACTING PROTEINS 1
TTF1	THYROID TRANSCRIPTION FACTOR 1
UBX	ULTRABITHORAX
WOX	WUSCHEL-related homeobox
WOX13~CHis	WOX13 fused to 6xHis tag at its C-terminus; other fusion proteins were named accordingly
WUS	WUSCHEL

1 Introduction

1.1 The homeodomain

1.1.1 Short survey of the homeodomain and classification by structural characteristics

In the mid 1980's a highly conserved region was identified in *Drosophila* genes capable of homeotic transformation of body segments: the discovery of the homeobox had taken place (McGinnis, Levine et al. 1984; Scott and Weiner 1984). The homeobox encodes a DNA binding domain, which was named "homeodomain" in accordance with the previous terminology. It consists of 60 amino acids, some of which are nearly invariant as e.g. the WFXN motif at the successive residues 48 to 51 and an arginine residue at position 5 (Fig. 1). Several homeodomain structures were determined utilizing nuclear magnetic resonance (NMR) spectroscopy and X-ray crystallography and the high degree of structural conservation is remarkable. The general homeodomain structure is composed of an N-terminal arm preceding three α -helices (helix I, II and III) (Fig. 2A and C) that are arranged to form a globular shaped protein with a conserved hydrophobic core ("h" above sequence logo in Fig. 1, Fig. 2B and C). Helix I, the most N-terminal helix, and helix II are connected by a loop and run antiparallel, whereas a turn places helix III roughly perpendicular to helix II, which thus form the well known helix-turn-helix (HTH) motif (Qian, Billeter et al. 1989).

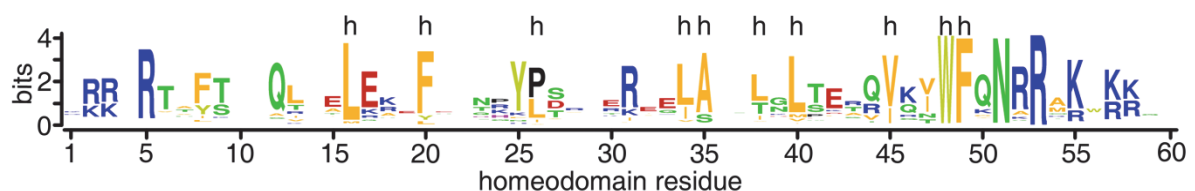


Fig. 1: Frequency of amino acids as function of animal homeodomain residue. h: conserved amino acid residues of the hydrophobic core modified after a review about the homeodomain (Bürglin 1994)

Soon after the discovery of the homeotic genes of *Drosophila* genes encoding homeodomains were detected in other species and not only in the animal kingdom but also in fungi and plants indicating an early evolution of this particular DNA binding domain (Schulz, Banuett et al. 1990; Vollbrecht, Veit et al. 1991; Schena and Davis 1992; Stankis, Specht et al. 1992). But not all of the newly discovered homeodomain sequences fitted perfectly into the canonical range of 60 amino acids (Hall and Johnson 1987). In general sequences, which require the insertion of gaps

or exclusion of residues for optimal alignment to the typical 60 amino acid homeodomains, are referred to as atypical homeodomains (reviewed in Bürglin 1994). A distinct subclass within the atypical homeodomains is formed by TALE homeodomains, the abbreviation of: “Three amino acids loop extension” (Bertolino, Reimund et al. 1995). As the

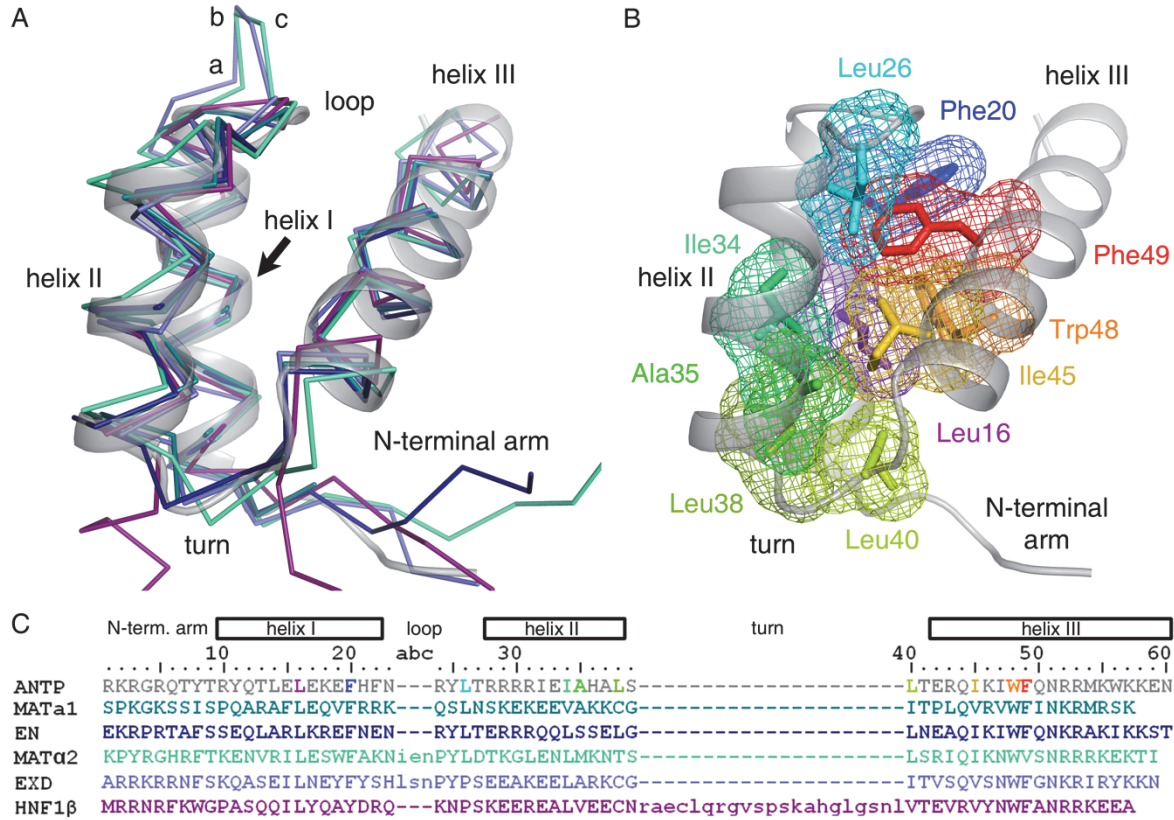


Fig. 2: General characteristics of the homeodomain structure. A) Superposed structures of ANTP HD (grey, PDB ID 9ANT), EN HD (dark blue, PDB ID 1ENH), MATA1 HD (teal, PDB ID 1YRN), EXD HD (slate blue, PDB ID 1B8I), MATA2 HD (turquoise, PDB ID 1APL) and HNF1β (purple PDB ID 2H8R); ANTP HD is displayed as ribbon, the other structures as lines connecting the Cα-atoms. B) ANTP HD shown as grey ribbon; highly conserved residues, which form the hydrophobic core, are depicted as coloured sticks surrounded by a mesh. C) Sequence alignment of homeodomains displayed in A and B; colouring is according to A and B; rectangles above sequence alignment indicate positions of helices (Fraenkel and Pabo 1998; Passner, Ryoo et al. 1999).

name implies those homeodomains are characterized by the integration of three additional residues that protrude out of the loop region into the solvent (Wolberger, Vershon et al. 1991) and are commonly referred to as residues a, b and c (Fig. 2A and C). Homeodomains harbouring this distinct insertion of three amino acids were identified not only in animals but also in plants and phylogenetic analysis suggests a common eukaryotic ancestry of all TALE homeodomains (Mukherjee, Brocchieri et al. 2009). Moreover, a domain that facilitates dimerization, the MEINOX domain, is shared in sequence and function between TALE homeodomain subclasses in animals - PBC and MEIS - and plants - KNOX and BEL (Bellaoui, Pidkowich et al. 2001;

Smith, Boschke et al. 2002; Mukherjee, Brocchieri et al. 2009). All these findings indicate a common origin of typical and TALE homeodomains in eukaryotes prior to the split of the major lineages of fungi, animals and plants.

Integration of additional amino acids is believed to mainly occur in between the helices, which would not perturb homeodomain structure and function (Bürglin 1994). This seems to be confirmed by structures of atypical homeodomains, where inserted amino acids are looped out from the globular homeodomain structure, e.g. in the TALE homeodomains of EXTRADENTICLE (EXD, PDB ID 1B8I) and MATING TYPE $\alpha 2$ (MAT $\alpha 2$, PDB ID 1APL) or in the atypical homeodomain of HEPATOCYTE NUCLEAR FACTOR 1 β (HNF1 β , PDB ID 2H8R) (Fig. 2A and C). In all cases the relative orientation of the helices to each other remains largely unaffected and thus aside from the additional residues structural deviation between typical and atypical homeodomains is low (Table 1).

Table 1: Root mean square deviation (RMSD) of homeodomains belonging to different structural classes

homeodomains		MATa1 HD	EXD HD	MAT $\alpha 2$ HD	HFN1 β HD
	PDB ID	(1YRN)	(1B8I)	(1APL)	(2H8R)
EN HD	(1ENH)	0.70	0.98	1.12	0.85
MATa1 HD	(1YRN)		0.71	1.05	0.77
EXD HD	(1B8I)			1.16	1.24
MAT $\alpha 2$ HD	(1APL)				1.39

The high structural conservation of yeast and animal homeodomains (Fig. 2, Table 1) and the common eukaryotic ancestry of animal and plant homeodomains (Mukherjee, Brocchieri et al. 2009) suggest that available data on homeodomain structure and function may be applied to plant homeodomains. However, research has been focused mainly on typical and TALE animal homeodomains resulting in the determination of only few structures of atypical that are not TALE homeodomains and not a single one of plant origin. Though it does seem likely that major structural changes would impair homeodomain function the available data on atypical and not-animal homeodomains is still limited. In particular the high diversity of atypical homeodomains is not reflected in the current structural data set.

1.1.2 General concept of homeodomain DNA recognition

Despite a high variety in amino acid sequence and length not only the homeodomain structure but also the manner of DNA binding seems remarkably well conserved. This allows the direct comparison of different homeodomains and the definition of general rules for homeodomain

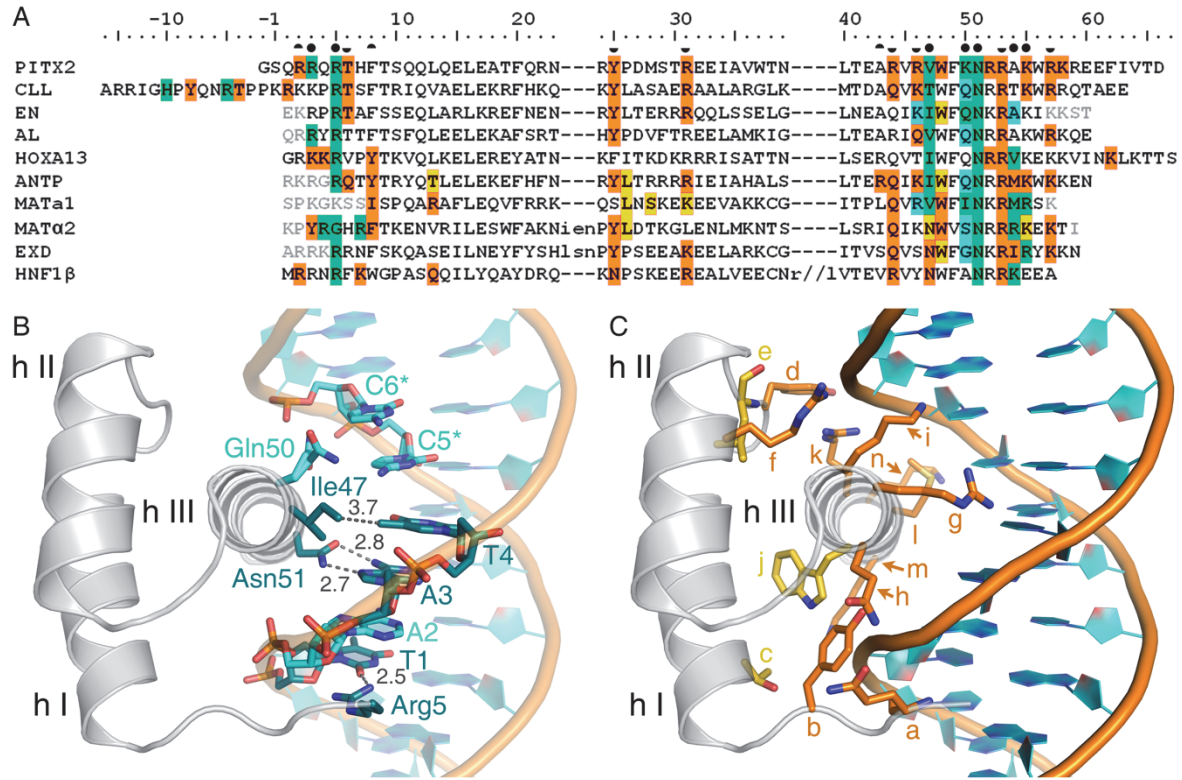


Fig. 3: Conserved homeodomain DNA recognition. A) Aligned homeodomain sequences compiled from a selection of solved homeodomain-DNA complexes: PITX2 HD (PDB ID 2LKX), CLL (PDB ID 3A01), EN HD (PDB ID 3HDD), AL HD (PDB ID 3A01), HOXA13 HD (PDB ID 2LD5), ANTP HD (PDB ID 9ANT), MATa1 HD (PDB ID 1YRN), MATa2 HD (PDB ID 1APL and 1YRN), EXD HD (PDB ID 1B8I) and HNF1β HD (PDB ID 2H8R); additional amino acids of atypical HD are shown in small letters; two slashes (/) depict cuts in sequence of HNF1β HD; grey coloured amino acids indicate unresolved homeodomain residues; amino acids with highlighted background, depict the DNA contacts of the homeodomain as described in the according publication: direct contacts, hydrogen bonds or hydrophobic interactions, with the nucleotide bases are coloured in teal, indirect, e.g. water mediated, base contacts in cyan, residues interacting directly or indirectly with the DNA backbone in orange or yellow, respectively; residues co-varying with DNA bases were identified by Noyes and colleagues (semi circle, top half) (Noyes, Christensen et al. 2008), Berger and colleagues (semi circle, bottom half) (Berger, Badis et al. 2008) or both (circle). B) and C) ANTP HD was selected as an example to visualize homeodomain DNA binding; side chains of residues contacting nucleotide bases (B) or the DNA backbone (C) are displayed as sticks; colouring of DNA contacting residues is according to (A); helix I, II and III are abbreviated as h I, h II and h III. B): contacts are depicted as grey dashed lines, distances are given in Ångström. C) DNA backbone contacts are established by Gln6 (a), Tyr8 (b), Thr13 (c), Tyr25 (d), Leu26 (e), Arg31 (f), Arg43 (g), Gln44 (h), Lys46 (i), Trp48 (j), Arg53 (k), Met54 (l), Lys55 (m) and Lys57 (n); Gln6 (a) and Leu26 (e) interact with the DNA backbone via their main chain.

DNA recognition and specificity (reviewed in Gehring, Qian et al. 1994; Berger, Badis et al. 2008; Noyes, Christensen et al. 2008). As can be observed in all available structures of homeodomain-DNA complexes, the N-terminal arm fits into the minor groove of the DNA and the third α -helix is inserted into the major groove, which is why helix III is also referred to as the DNA recognition helix (Fig. 3B, C). Essentially the same residues compose the interaction

surface of every homeodomain-DNA complex. Depending on their location and identity of the individual amino acid residues of the interaction surface either aid in the correct orientation of the DNA helix by contacting its sugar-phosphate backbone (orange, yellow Fig. 3A, C) or facilitate sequence recognition by interacting with the nucleotide bases (teal, cyan Fig. 3A, B). In the following these contacts will be described in more detail with residue numbering according to the position in typical homeodomains.

Highly conserved contacts between the homeodomain and the DNA backbone are provided by the residues at positions 25 and 31, which are located in the loop and helix II, as well as 44, 46, 48, 53 and 57 in helix III. In case of the N-terminal arm the situation appears to be less predetermined, which may be due to its comparably high structural flexibility. Still residues 6 or 8 show some degree of conservation in contacting the sugar-phosphate backbone. In general the nucleotide bases are contacted by amino acids at position 5 in the N-terminal arm and 47, 50, 51 and 54 in the recognition helix, whose side chains protrude into the minor and major groove, respectively. It has been reported that also residues 3, 4, 7 and 55 may participate in DNA sequence recognition, however, it seems only provided that those positions are occupied by an arginine as in cases of the homeodomains of ARISTALESS (AL) and PITUITARY HOMEODOMAIN PROTEIN 2 (PITX2) or EXD and

MATa1 (Fig. 3A) (Li, Stark et al. 1995; Passner, Ryoo et al. 1999; Chaney, Clark-Baldwin et al. 2005; Miyazono, Zhi et al. 2010) Two contacts between the highly conserved homeodomain residues Arg5 and Asn51 (Fig. 1) and specific nucleotides (Fig. 3; Table 2) allow the normalization of each DNA binding sequence to those reference points. The nature of these contacts is discussed in more detail in section 1.1.3.1 and 1.1.3.2. Here they are mentioned briefly to outline the labelling, which is used to refer to a specific nucleotide position in this thesis. The invariant residue Asn51 of the recognition helix specifies an adenine in the DNA sequence (Fig. 3B; Table 2). The thus conserved adenine is referred to as position 3 of the DNA sequence (A3) in this thesis. Separated by one nucleotide, position 1 of the DNA sequence is contacted by Arg5, which determines a thymine in most cases (Fig. 3B; Table 2). The DNA strand harbouring A3 is referred to as the sense DNA strand from hereon. Nucleotides of the complementary or anti-sense strand are indicated by an asterisk (*) but retain position numbers of the sense strand, e.g. T3*. For the sake of clarity designations of other publications are adjusted accordingly in this thesis.

1.1.3 Functional specificity of homeodomain monomers

1.1.3.1 The sequence specificity is largely defined by the amino acid composition of residues with direct contact to nucleotide bases

The highly defined structure and conserved DNA binding mechanism allow the direct comparison of different homeodomains and thus the identification of residues, which confer sequence specificity in this family of transcription factors. This forms the basis for two large scale experiments concerned with the covariance of amino acid sequence and preferred DNA recognition sequence of mouse and *Drosophila* homeodomains (Berger, Badis et al. 2008; Noyes, Christensen et al. 2008). Utilizing different experimental approaches both studies essentially identified the same residues to be crucial for DNA recognition and specificity. It is not particularly surprising that all of them are part of the homeodomain-DNA contact surface (Fig. 3A, circles and semi circles above sequence alignment). Those results are consistent with previous findings based on structural analysis of several homeodomain - DNA complexes (summarized and colour coded in Fig. 3A) and various mutational experiments, which will be discussed in more detail below.

Many of the known DNA recognition sequences preferred by homeodomains comprise a six base pair element rich in thymine and adenine, which often forms a TAAT-core motif at positions 1 to 4 or the innate sequence TGAT (Gehring, Affolter et al. 1994; Wilson, Sheng et al. 1996; Banerjee-Basu, Moreland et al. 2003). As one would expect, residues contacting bases of the frequently occurring core motifs are rather conserved themselves (Wilson, Sheng et al. 1996). Probably the most striking example is Asn51 of the homeodomain motif in helix III (WFxN). By establishing a very specific bidentate contact, the invariant Asn51 defines the conserved A3 mentioned in the previous section (Li, Stark et al. 1995; Fraenkel and Pabo 1998; Fraenkel, Rould et al. 1998; Passner, Ryoo et al. 1999; Chaney, Clark-Baldwin et al. 2005; Miyazono, Zhi et al. 2010; Zhang, Larsen et al. 2011) (Fig. 3B and Table 2). Though not as conserved but still apparent in the majority of the homeodomains, an arginine at position 5 (Arg5) contacts the DNA via the minor groove and interacts with the first nucleotide of the recognition sequence and sometimes even position -1 (Bürglin 1994; Fraenkel and Pabo 1998; Fraenkel, Rould et al. 1998; Passner, Ryoo et al. 1999; Miyazono, Zhi et al. 2010; Zhang, Larsen et al. 2011) (Table 2). The occurrence of a thymine at position 1 (T1) of the preferred DNA binding sequence highly correlates with an arginine at residue 5 of the homeodomain (Noyes, Christensen et al. 2008). At first glance this specificity is sort of peculiar, because viewed from the minor groove the two

pyrimidine nucleotides, thymine and cytosine, are the same. Thus Arg5 should not be able to distinguish between these two. Curiously the homeodomains of ENGRAILED (EN) and THYROID TRANSCRIPTION FACTOR 1 (TTF1) both contain an Arg5 but differ in their specificities for position 1. EN HD binds sequences with a thymine and TTF1 HD with a cytosine at position 1 but neither would recognize each others sequences (Ades and Sauer 1995; Damante, Pellizzari et al. 1996). Indeed sequence specificity conferred by the N-terminal arm seems to depend on a combination of a variety of different factors, which will be discussed in more detail in the next section.

The other nucleotides of the DNA binding sequence, positions 2 and 4 of the TAAT-core as well as positions outside the core, seem to be determined by the integrated biochemical properties of several, sometimes competing residues. Position 2 of the sequence can be contacted by residues located in the N-terminal arm via the minor groove as well as by residues of the recognition helix via the major groove. According to Noyes and colleagues the identity of the nucleotide at position 2 is specified by argenines at positions 2, 3 or 55 (Noyes, Christensen et al. 2008). The results provided by Noyes and colleagues suggest, that an arginine at either residue 2 or 3 has a similar effect on the preferred binding sequence. Other studies, however, indicate slight differences depending on the context of other residues in the N-terminal arm, which will be discussed in more detail in section 1.1.3.2.

Aside of these small divergences, findings concerned with sequence specificity conferred by Arg3 and Arg55 are consistent in several studies. Arg55 defines a guanine at position 2 (G2), whereas Arg3 interacts with a thymine of the complementary DNA strand, which thus results in an adenine at position 2 (A2) (Li, Stark et al. 1995; Passner, Ryoo et al. 1999; Chaney, Clark-Baldwin et al. 2005; Miyazono, Zhi et al. 2010) (Table 2). So far homeodomains, which either contain Arg55 or Arg2/3, have been the subject of research and no data was found on sequence preferences given that Arg2/3 and Arg55 are present simultaneously.

Table 2: Interaction of homeodomain residues and DNA bases compiled from a selection of published complexes: ANTP (Fraenkel and Pabo 1998), AL and CLL (Miyazono, Zhi et al. 2010), EN (Fraenkel, Rould et al. 1998), EXD (Passner, Ryoo et al. 1999), HNF1 β (Lu, Rha et al. 2007), HOXA13 (Zhang, Larsen et al. 2011), MATa1 and MATa2 (Wolberger, Vershon et al. 1991; Li, Stark et al. 1995), PITX2 (Chaney, Clark-Baldwin et al. 2005); top row: nucleotide position relative to A3, which is defined by its conserved interaction with Asn51 of the WFXN-motif in helix III; in each row the residues which contact the respective nucleotide are listed - nucleotides of the complementary strand are indicated by an asterisk (*) – below the bound DNA sequence is shown; nucleotides in brackets are not contacted by any homeodomain residue in the structure; residue labels are normalized to their corresponding position in typical homeodomains.

	-1	1	2	A3	4	5	6	7
ANTP		Arg5: T		Asn51: A Ile47: A	Ile47: T	Gln50: C*	Gln50: C*	
AL	Arg5: A*	T Arg5: T	(A) Arg3: T* Asn51: A	A Asn51: A Val47: A	T Val47: T	G	G Gln50: T*	
CLL	T Arg5: A*	T Arg5: T	A Asn51: A	A Asn51: A Thr47: A	T Arg-5: A* Thr47: T	(T) Arg-5: A*	A His-10: G Gln50: C*	(A)
EN	T	T Arg5: T	A	A Asn51: A Ile47: A	T Ile47: T	T	G Gln50: T*	(C)
EXD		T Arg5: T	(A) Arg55: G	A Asn51: A	T Asn51: A*	(T) Gly50: A*	A	(C)
HNF1 β	(C)	T Arg5: T	G	A Asn51: A	T Lys54: G*	T	(T)	(A)
HOXA13	Arg5: A*	T Val54: T	(C, A) Val54: T Asn51: T	A Asn51: A Val54: T*	C Ile47: T	Ile47: T		
MATa1	T	T	T Arg55: G	A Asn51: A Arg55: T* Val47: A	T	T Met54: C*	(T)	(T) Ile50: T*
MATa2	(A) Arg7: A	(T)	G Arg4: A* Gly5: T	A Asn51: A Arg4: A	(T) Arg54: G*	G Ser50: T*	(T) Ser50: A*	A
PITX2	A Arg5: G*	(T)	T Arg3: A Val47: A	A Asn51: A Val47: A	C Val47: T	A Lys50: G*	T Lys50: G*	(G)
	C	(T)	A	A	T	C	C	(C)

Suggested by biochemical studies concerned with DNA specificity and affinity of an EN mutant homeodomain, A2 may be contacted via the major groove by residues other than Arg55 (Ades and Sauer 1995). Asn51 is the most likely candidate as it was shown to contact A2 or T2 in some homeodomain structures (Table 2) (Miyazono, Zhi et al. 2010; Zhang, Larsen et al. 2011). However the interactions appear rather unspecific and thus Asn51 is unlikely to be a strong specificity determinant for position 2. Less frequent and thus probably rarely mentioned is a putative influence of Val54 on the identity of the nucleotide at position 2. Apparent in the structure of the HOMEODOMAIN A 13 homeodomain (HOXA13 HD) in complex with its high

affinity binding site Val54 defines a thymine at position 2 (T2) and even seems able to overrule the sequence preference of Arg2, which is present simultaneously (Knosp, Saneyoshi et al. 2007; Zhang, Larsen et al. 2011) (Table 2).

The positions 4, 5 and 6 of the DNA recognition sequence seem to be specified mainly by combinatorial effects of residues 47, 50 and 54 in helix III (Noyes, Christensen et al. 2008). Their deterministic power is highly dependent on the identity of the individual amino acids and seems to be exerted in a hierarchical order for each of the three DNA positions. The ubiquitous thymine at position 4 (T4) of the TAAT-core is defined by the two most common amino acids at residue 47, isoleucine and valine (Ile/Val47), which both are able to interact with the methyl group of thymine (Fig. 3B) (Bürglin 1994; Fraenkel, Rould et al. 1998; Chaney, Clark-Baldwin et al. 2005). Only the influence of Tyr54 seems to be stronger than that of Ile/Val47. The occurrence of Tyr54 coincides with a guanine at position 4 (G4) apparently regardless of the identity of the other residues and also when co-occurring with Ile/Val47 as e.g. in case of TTF1 HD (Noyes, Christensen et al. 2008; Chu, Noyes et al. 2012). Analysis of binding preferences reveals that wild type TTF1 HD recognizes sequences containing G4, whereas upon mutation of Tyr54 to methionine the homeodomain specificity is switched to T4, which is likely to represent the now unblocked deterministic power of Ile47 (Damante, Pellizzari et al. 1996). Consistent with this finding, mutation of Met54 in the homeodomains of ANTP and Ala54 in EN to tyrosine leads to preferential binding of G4 instead of the wild type T4 (Damante, Pellizzari et al. 1996). Thus Tyr54 appears to be highest in ranking to define the nucleotide at position 4 of the DNA recognition sequence. However it is not a very common residue in homeodomains, which might be the reason why the more frequent Ile/Val47 is able to specify T4 of the omnipresent TAAT-core most of the time. The influence of residue 50 on position 4 is subtler and seems to mainly lie in the blockage of Arg54. If residue 47 is not occupied by an isoleucine or valine, Arg54 may determine a cytosine at position 4 (C4) by binding to the guanine of the complementary DNA strand (G4*) as e.g. in case of the MATING TYPE α 2 homeodomain (MAT α 2 HD; PDB ID 1APL) (Table 2) (Li, Stark et al. 1995; Noyes, Christensen et al. 2008). However analysis of the binding sequences of mutated BICOID (BCD) and EN homeodomains suggest that the mutation of Ile47 in context of Arg54 is not always sufficient to switch binding preferences from T4 to C4 (Noyes, Christensen et al. 2008). Rather Lys50 and possibly also Gln50 seem to interfere with Arg54 specification of position 4 and only additional mutation of residue 50 to alanine allows Arg54 to exploit its full potential.

Other polar amino acids, like threonine or asparagine, at residue 47 also seem to prefer thymine at position 4 (Noyes, Christensen et al. 2008). However, the influence in particular of Asn47 seems to be comparably low. First, apparent in several structures, Asn47 seems to mainly interact with the DNA backbone (Li, Stark et al. 1995; Passner, Ryoo et al. 1999; Lu, Rha et al. 2007) (Fig. 3A). Furthermore the preferences of Asn47 are overruled easily. For example in the homeodomain of the HEPATOCYTE NUCLEAR FACTOR 1 β (HNF1 β ; PDB ID 2H8R) Lys54 recognizes a guanine of the complementary DNA strand (G4*) and thus annuls the preference of the co-occurring Asn47 for A4* (Lu, Rha et al. 2007). Unlike Asn47, Thr47 seems able to directly interact with T4 via its methyl group, as can be seen in the structure of the CLAWLESS homeodomain (CLL HD, PDB ID 3A01) (Miyazono, Zhi et al. 2010). According to Noyes and colleagues the strength of this interaction additionally depends on the identity of the amino acid at residue 43 and is stronger, if it is occupied by either valine or threonine (Noyes, Christensen et al. 2008). This finding indicates the putative formation of a hydrophobic network, which might support the specification of T4 by Thr43. Other studies, however, fail to detect an interdependence of residue 43 and the preferred DNA binding sequence (Berger, Badis et al. 2008; Chu, Noyes et al. 2012). Additionally the side chain of threonine is very short so one can imagine that the hydrophobic interaction of its methyl group with that of T4 could be easily interrupted. Taken together the influence of Tyr54, Ile/Val47 and Arg54 on position 4 of the DNA binding sequence is well studied and presumably the identity of position 4 is predictable if either one of those is present in the homeodomain. But polar amino acids at residue 47 seem to have only a weak preference for T4, which may be modulated or overpowered easily by several other factors.

The DNA positions 5 and 6 lie outside the rather conserved TAAT-core and were believed to mainly account for differential DNA binding of HOX homeodomains. Several studies were concerned with their specification and in the course of the according experiments a single amino acid at residue 50 of the homeodomain emerged as the main determinant. Single mutation of residue 50 was sufficient to alter or interchange specificity for positions 5 and 6 of various homeodomains containing glutamine, lysine or serine at residue 50 (Hanes and Brent 1989; Treisman, Gonczy et al. 1989; Damante, Fabbro et al. 1994; Wilson, Sheng et al. 1996; Tucker-Kellogg, Rould et al. 1997). Apparent in the NMR structures of ANTP HD (PDB ID 1HOM) and PITX2 HD (PDB ID 2LKX) the side chain of Gln50 and Lys50, respectively, establishes fluctuating contacts with the two nucleotides adjacent to the TAAT-core, which might allow the simultaneous specification of two DNA positions by one residue (Billeter, Qian et al. 1993;

Chaney, Clark-Baldwin et al. 2005). However some results raised questions as to how definite the specification is especially in case of Gln50. The mutated EN HD, in which Gln50 is replaced by alanine, displays only slightly decreased affinity and specificity compared to wild type (Ades and Sauer 1994). Structural analysis revealed that the DNA contact of the Gln50 side chain is replaced by highly coordinated water molecules in the Ala50 mutant of EN HD (Grant, Rould et al. 2000). These findings prompted the conclusion, that the influence of Gln50 on the preferred binding sequence is rather low. Deduced from the catalogue of specificity determinants established by Noyes and colleagues and the fast amount of mutational and biochemical studies, residue 50 appears to be the central player in defining positions 5 and 6, nonetheless (Noyes, Christensen et al. 2008). But possibly not the single player as residues 47 and 54 were shown to be associated with the specification of those positions as well (Noyes, Christensen et al. 2008; Chu, Noyes et al. 2012). In the folded protein their side chains frame the one of residue 50 and might be able to modulate or even interfere with its contact to the DNA. In extant homeodomains the occurrence of Gln50 coincides with various different combinations of amino acids at residue 54, whereas Lys50 and Cys50 seem to be very restrictive (Damante, Pellizzari et al. 1996). In particular Lys50 and Met54, a pair not observed in native homeodomains so far, was demonstrated to clearly interfere with each other (Pellizzari, Tell et al. 1997). Mutations to obtain Lys50 and Met54 in the same homeodomain led to weakened DNA binding activity and impaired base discrimination (Damante, Pellizzari et al. 1996; Pellizzari, Tell et al. 1997). Taken together Gln50 may be more susceptible to the external influence of neighbouring side chains than for example Lys50. But on the other hand Gln50 seems to allow a greater variety of accompanying amino acids, which in turn modulate sequence specificity of Gln50. So rather than playing only “a modest role in DNA recognition” as has been suggested by Grant and colleagues (Grant, Rould et al. 2000), Gln50 might be very important for the diversity of DNA binding sequences, which also may explain why glutamine is by far the most common amino acid at residue 50 (Fig. 1) (Gehring, Affolter et al. 1994; Banerjee-Basu, Moreland et al. 2003).

1.1.3.2 Sequence recognition of the N-terminal arm

The most frequent contacts of residues located in the N-terminal arm to nucleotide bases are established between the residues Arg2, Arg3 and Arg5 and positions 2 and 1 in the minor groove of the preferred DNA recognition sequence (Fig. 1 and Fig. 3) (Gehring, Affolter et al. 1994). But at least in case of MAT α 2 HD Arg4 and Arg7 seem to have the potential to interact with nucleotide bases, as well, which then define position 2 and -1, respectively (Li, Stark et al.

1995). This creates the impression that possibly arginine is the only amino acid, which fulfils the requirements for contacting nucleotide bases via the minor groove of the DNA structure. A possible explanation has been given by the findings of Joshi and colleagues, who determined the structures of a heterodimer consisting of the homeodomains of SEX COMP REDUCED (SCR) and its cofactor EXTRADENTICLE (EXD) bound cooperatively to two different DNA sequences (Joshi, Passner et al. 2007). One of the sequences is a SCR-specific and the other a consensus HOX-EXD binding site. Their comparative analysis of the two structures and additional experiments indicate that the N-terminal arm of SCR HD distinguishes between the DNA-shape of the two different sequences. The SCR-specific site is characterized by two minima of negative electrostatic potential occurring at narrow minor groove widths, which are recognized by the basic residues of the N-terminal arm. A secondary analysis of the results of two high throughput experiments conducted with mouse (Berger, Badis et al. 2008) and *Drosophila* (Noyes, Christensen et al. 2008) homeodomains suggest a high dependency of basic amino acids at residues 2, 3 and 5 and minor groove width at position 1 and 2 of the DNA binding site (Dror, Zhou et al. 2014). Also biochemical analysis of the DNA binding properties of an EN mutant homeodomain suggests a coupling of the DNA positions 1 and 2 (Ades and Sauer 1995). Possibly the identity of the nucleotides at those two positions determines the minor groove width, which is in turn recognized by the N-terminal arm.

Besides the basic amino acids of the N-terminal arm, residues without physical contact to nucleotide bases or sometimes even the DNA backbone were shown repeatedly to have high influence of the DNA sequence recognition. Following the principle of “all or nothing” residue 8 plays a major role. At that position of the N-terminal arm phenylalanine and tyrosine are considerably well conserved (Fig. 1) and mutation of Phe8 in MAT α 2 to alanine results in a significantly reduced repressive activity of the MAT α 1-MAT α 2 and MCM1-MAT α 2 dimer (Mathias, Zhong et al. 2001). Members of the IROQUOIS-group are rare examples of homeodomains containing a native Ala8. As was reported by Noyes and colleagues, the IROQUOIS homeodomains show little base discrimination for positions 1 and 2 of the DNA binding site even though they contain Arg5 and Arg55, which are supposed to be strong specificity determinants for T1 and G2, respectively (Noyes, Christensen et al. 2008) (see also previous section). In structures of other homeodomains the side chain of Phe8 is inserted into a hydrophobic pocket between helix I and II. This prompted the conclusion that a large hydrophobic amino acid at residue 8 possibly stabilizes the flexible region of the homeodomain and thus might be required to facilitate the interaction of the N-terminal arm with the

DNA (Noyes, Christensen et al. 2008). The theory was tested utilizing the CAUP homeodomain as representative of the IROQUOIS-group. Mutation of Ala8 to phenylalanine elicited the selection of the expected sequence with T1 and G2. The effects of Ala8 on conformation of the N-terminal arm are possibly similar to those observed for Ile8 in the homeodomain of MATA1 (PDB ID 1YRN). Resolution of the N-terminal arm is low, which indicates a high flexibility of this part of the structure, and no interaction with the DNA could be detected for any of the residues 1 to 7 (Li, Stark et al. 1995).

An interesting example is also provided by the homeodomains of the *HOM*-genes of *Drosophila*, which determine segment identity along the anterior-posterior body axis. In many cases the functional divergence could be mapped to amino acids of the N-terminal arm. The homeodomains of ANTENNAPEDIA (ANTP) and SEX COMB REDUCED (SCR) differ only at the homeodomain positions 1, 4, 6 and 7 from each other, none of which interact with nucleotide bases. However mutation of those few residues in SCR is sufficient to confer ANTP-like segment identity (Furukubo-Tokunaga, Flister et al. 1993). Likewise amino acids at positions 3, 6 and 7 of the N-terminal arm account for diverging sequence preferences of ULTRABITHORAX (UBX) and ABDOMINAL-B (ABD-B) (Egger, Jackson et al. 1994). The DNA sequence preferences of ABD-B seem to differ greatly from the homeodomains of the other *HOM*-genes. According to the expression of the *HOM*-genes along the anterior-posterior body axis, Noyes and colleagues observed an increasing tolerance of the respective homeodomain for thymine at position 2 (T2) of the DNA recognition sequence (Noyes, Christensen et al. 2008). Eventually the most posterior one, ABD-B HD, does indeed prefer thymine over the more common adenine and thus forms its own sequence specificity group. This gradual change in binding preferences seems to correlate with the exchange of the base contacting residues Lys2/Arg3 to Arg2/Lys3. But single mutation of Arg3 to lysine is not sufficient to alter UBX HD DNA binding specificity (Egger, Jackson et al. 1994). Additionally changes of residues 6 and 7 are required to obtain DNA sequence preferences resembling that of ABD-B HD.

Even more extreme is the sequence specificity conferred by residues 6, 7 and 8 in case of the homeodomain of the THYROID TRANSCRIPTION FACTOR 1 (TTF1). Similarly to the majority of homeodomains TTF1 HD contains an arginine at residue 5 (Arg5), which seems highly correlated with a preference for thymine at position 1 (T1) (Table 2) (Noyes, Christensen et al. 2008). TTF1 HD, however, recognizes sequences with a cytosine at position 1 (C1) (Damante, Fabbro et al. 1994). Mutation of the Val6, Leu7 and Phe8 in TTF1 HD to the

corresponding amino acids apparent in ANTP HD - Gln6, Thr7, Tyr8 - shifted the preference of TTF1 HD to T1 containing sequences (Damante, Pellizzari et al. 1996). Vice versa the accordingly mutated ANTP HD recognized C1 sequences. Consistent with these results the two studies, which discuss shape dependent DNA recognition by the N-terminal arm, identified residues other than the usual suspects, residues 2, 3 and 5, to influence DNA binding. Joshi and colleagues observed a loss of specificity after mutating Gln4 in SCR to glycine (Joshi, Passner et al. 2007). Also Dror and colleagues detected covariation between residues 4, 6 and 7 and the ability of a homeodomain to recognize the DNA shape, which apparently is independent of a specific identity of the residues, like e.g. arginine (Dror, Zhou et al. 2014). This excludes the possibility of direct physical interaction with an increased negative electrostatic potential of the minor groove. Likewise Gln4 in SCR does not contact nucleotide bases (PDB ID 2R5Z).

Based on their own structures or on published structures of other homeodomains all studies concluded similarly that residues of the N-terminal arm without direct base contact seem to influence DNA binding by determining the conformation of the N-terminal arm and thus its position over the minor groove. Taken together these results suggest that DNA recognition and sequence specificity of the N-terminal arm is determined by a combination of structural conformation defined commonly by residues 4, 6, 7 and 8 as well as base specific contacts and/or shape readout by basic amino acids most frequently found at residues 2, 3 and 5 (Fig. 1).

1.1.3.3 Influence of backbone contacts outside the N-terminal arm on DNA recognition

The interaction of the homeodomain with the sugar-phosphate backbone has been reported repeatedly to influence DNA affinity, specificity and *in vivo* function. Substitutions of residues contacting the DNA backbone in MAT α 2 HD for alanine result in a significantly reduced repressive activity of the MAT α 2-MCM1 complex (Mathias, Zhong et al. 2001). A covariation analysis of mouse homeodomain residues and their respective preferred DNA sequence also identified residues, which establish conserved contacts with only DNA backbone (Berger, Badis et al. 2008). The homeodomain positions, which typically contact the DNA backbone, comprise residue 25 in the loop, 31 and - though more rarely - 28 at the N-terminus of helix II and the residues 44, 46, 48, 53 and 57 in the recognition helix (Fig. 3A and C) (reviewed in Gehring, Qian et al. 1994). Given the biochemical nature of the DNA backbone, it seems only consequential that the affinity of a homeodomain to its target sequence is at least partially determined by charged residues at the contact surface. Several mutational analyses of

homeodomains emphasize the importance of these residues for DNA recognition and *in vivo* function.

Substitution of Tyr25 with alanine in EN HD greatly affects its DNA binding capability *in vitro* (Sato, Simon et al. 2004). In case of MAT α 2 this mutation results in a massive reduction of MAT α 2-MCM1 mediated repressive activity *in vivo* (Mathias, Zhong et al. 2001). In the homeodomain of FUSHI TARAZU (FTZ) mutation of Tyr25 to isoleucine lowered its activity in a transactivation assay with cultured *Drosophila* cells to 33% of wild type level (Furukubo-Tokunaga, Muller et al. 1992). Substitution of Tyr25 with a short hydrophobic amino acid appears to be highly detrimental in many cases. The extent of allowed variation at residue 25, however, appears to vary between different homeodomains. Whilst EN HD seems to partly tolerate phenylalanine at residue 25 instead of tyrosine (Sato, Simon et al. 2004), the very same mutation almost completely abolishes DNA binding of PBX1 HD (Lu and Kamps 1996). Despite slight discrepancies the overall conclusion of all experiments is similar: Tyr25 is crucial for DNA recognition and the high degree of conservation throughout evolution (Fig. 1) indicates an important role of Tyr25 in homeodomain function that possibly can not be resumed by another amino acid easily.

Following the peptide chain towards the C-terminus of the homeodomain the next region, identified to be important for DNA-binding is separated by just a few residues. At the N-terminus of helix II quite a few homeodomains harbour a stretch of up to four consecutive basic amino acids ranging from residue 28 to 31. Since the N-terminus of helix II is in close proximity to the phosphate backbone in solved structures (Fig. 3C), it has been suggested that it may directly influence the affinity of homeodomains to DNA (Passner, Ryoo et al. 1999). The arginine stretch at residues 28 to 31 was demonstrated to be crucial for *in vivo* function of FUSHI TARAZU (FTZ) (Furukubo-Tokunaga, Muller et al. 1992). In particular the importance of Arg28 and 31, whose side chains direct towards the DNA interface, was emphasized by subsequent transactivation assays in cultured *Drosophila* cells in the same study. Arg31 establishes a rather highly conserved contact with the DNA backbone (Fig. 3A and C). Two experiments scanning for residues in mouse homeodomains and EN HD, which are involved in DNA recognition identify residue 31 as fundamental (Sato, Simon et al. 2004; Berger, Badis et al. 2008). But also the fact that mutations of Arg31 in homeodomains are connected to several human diseases corroborates the importance of this interaction site (D'Elia, Tell et al. 2001). In contrast to residue 31, the backbone contact established by residue 28 is less conserved and thus its function less well described. But besides the already mentioned pivotal role of residue 28 for

in vivo functionality of FTZ another study attained intriguing effects by its mutation in the homeodomain of PBX1 (Pre-B-cell leukaemia transcription factor 1). In a strive to assess structural determinants of cooperative DNA binding of PBX1 HD, mutation of Glu28 to argenine was shown to increase the affinity of monomeric PBX1 HD to DNA but concurrently to completely abolish cooperative DNA binding with HOXA5 (Lu and Kamps 1996).

Continuing along the protein chain the next residues establishing contact with the DNA backbone are already part of the recognition helix. Due to the deep insertion of helix III in the major groove almost all of its residues are located at the homeodomain – DNA interface. Thus nearly the entire helix III is involved in different aspects of DNA binding (Fig. 3A). Alanine substitution of either of the residues that interact with the sugar-phosphate backbone in MAT α 2 HD greatly reduces its transcriptional activity in complex with MCM1; even water mediated contacts, for example the one established by Asn47, are indispensable for MAT α 2 HD function (Mathias, Zhong et al. 2001). The few amino acids that face away from the DNA are the conserved hydrophobic core residues 45 and 49 (reviewed in Bürglin 1994) as well as residues 52 and 56, whose side chains project into the solvent. In some cases even the side chain of Arg52 bends downward and around helix III to reach the DNA backbone (Fig. 3A; PDB IDs 2LKX and 2LD5). In PBX1 HD and possibly in the whole PBC family, residues 52 and 56 interact with a forth helix that is formed C-terminally to the homeodomain and seems to stabilize the protein-DNA complex (Piper, Batchelor et al. 1999). Together with the homeodomain the forth helix comprise the minimum protein required to achieve DNA binding activity resembling that of full length PBX1 (Lu, Rha et al. 2007). Apparent in many other structures of homeodomains, e.g. BCD HD (PDB ID 1ZQR), a basic amino acid at residue 52 and an acidic amino acid at residue 17 establish a salt bridge and link helix I with helix III (Clarke 1995). In EN HD both residues are basic. Interestingly an acidic amino acid at residue 17 is structurally favoured, but the according mutant EN HD exhibits a dramatically decreased affinity to its preferred DNA binding sequence (Stollar, Mayor et al. 2003; Sato, Simon et al. 2004).

Two residues of helix III, which have not been observed to establish direct base contact, seem to indirectly influence the preferred DNA recognition sequence. Residue 46 seems able to modulate DNA binding of residue 50 (Mahony, Auron et al. 2007) and residue 43 appears to alter the specificity of Thr47 (Noyes, Christensen et al. 2008). In homeodomain structures the two residues of each pair are located in close proximity to each other. This prompted both studies to conclude that residues 43 and 46 may influence the conformation of the base contacting side chains of Thr47 and residue 50, respectively, which thus may affect their sequence preferences.

However in an analysis of the binding specificities of EN HD with randomly mutated residues 43, 46, 47, 50, 54, no significant covariation could be detected for residues 43 and 46 and any position of the DNA sequence (Chu, Noyes et al. 2012). Possibly the effect of the modulating residues is rather subtle and may depend strongly on the amino acid at the residues with the actual base contact, as has been suggested for the pair 43-47, where only the sequence preference of threonine at residue 47 was altered (Noyes, Christensen et al. 2008). That might be a reason why their influence on DNA binding is difficult to detect in large scale experiments, whereas the importance in a subdivision or individual homeodomains may be more apparent. In a complementation assay of *ftz* mutant flies with FTZ-MSH chimeras (FUSHI TARAZU and MUSCLE SEGMENT HOMEODOMAIN) residue 43 in combination with the previously discussed arginine stretch at the N-terminus of helix II were demonstrated to be crucial for effective rescue (Furukubo-Tokunaga, Muller et al. 1992).

1.2 The WUSCHEL-related homeobox (WOX) family

1.2.1 Short introduction on the evolution of terrestrial plants

Land plants originated from unicellular green algae that, in contrast to their marine sister group inhabited fresh water and subsequently colonized the land about 432 to 476 million years ago (reviewed in Leliaert, Verbruggen et al. 2011). Bryophytes, which include hornworts, liverworts and mosses, constitute the group of the oldest terrestrial plant species (Wickett, Mirarab et al. 2014). Though some parts of the bryophyte plant body are reminiscent of tissues and organs in derived plant species, the body plan seems more primitive and specialized tissues are rare. The acquisition of water, for example, is accomplished by the whole plant body, whereas a rhizoid, which reminds one of roots in higher plants, primarily serves as fixation to the ground. However as the evolution of land plants continued the plant body became more complex and division of labour increased. The development of specialized organs and tissues in higher plants facilitated the invasion of new ecological niches. A plant vasculature, one of the first major structural innovations, is first detected in species belonging to the sister group of bryophytes, which was therefore termed vascular plants (tracheophytes). Lycophytes, e.g. *Selaginella moellendorffii*, and ferns (monilophytes) are basal representatives of the group of vascular plants (Wickett, Mirarab et al. 2014). The subsequent appearance of seeds marks the emergence and rapid radiation of seed plants (spermatophytes), which split into two large groups, the flowering plants (angiosperms) and gymnosperms (Chaw, Parkinson et al. 2000). Cycads and ginkgoales are basal

gymnosperm groups, whereas conifers and gnetales relate to the same group but developed later during evolution (Chaw, Parkinson et al. 2000). The most basal branch of angiosperms is composed of Amborellales, with *Amborella trichopoda* being its sole extant representative. Water lilies (Nymphales) and Magnoliids constitute other basal angiosperm branches, which evolved prior to the split into two major groups, the eudicots and the monocots (Soltis, Bell et al. 2008). The latter include grasses amongst others, e.g. important crops such as barley, rice and maize, whereas tomato or tobacco and also the often mentioned model organism *Arabidopsis thaliana* belongs to the group of eudicots.

1.2.2 Discovery of a novel homeodomain family, classification and evolution of its members

Plants and plant organs grow from so called meristems, which harbour a small population of pluripotent cells. Following germination the activity of two embryonic meristems at either tip of the main growing axis, the root and shoot apical meristems (RAM and SAM) give rise to all below and above ground parts of the postembryonic plant body of *A. thaliana*. The floral meristems (FM) develop from apical meristems after switching from the vegetative phase to reproductive phase and terminate in flowers. In order to identify novel players involved in SAM development and maintenance of plant stem cells, *A. thaliana* seeds were randomly mutagenized and subsequently the seedlings screened for a defective meristem in the shoot apex. (Laux, Mayer et al. 1996). One gene identified was named *WUSCHEL* (*WUS*) and plants exhibiting the *wus-1* mutation display impaired plant growth and defects in flower development resulting in sterile plants of reduced height. The conspicuous phenotype prompted several other experiments, which showed that the stem cells of SAM and FM differentiate in mutant plants leading to the premature termination of affected meristems (Mayer, Schoof et al. 1998; Stuurman, Jäggi et al. 2002; Kieffer, Stern et al. 2006). Thus *WUS* activity in *A. thaliana* and its homologues in other eudicot species is required for the maintenance of a population of pluripotent cells in SAM and FM.

In addition to the discovery that *WUS* is involved in the regulation of stem cell homeostasis in plants, the gene encodes a characteristic homeodomain sequence distinct from the ones previously described (Mayer, Schoof et al. 1998), which encouraged further research. Several other genes were detected encoding for a homeodomain similar to that of *WUS* and thus were named *WUSCHEL-related homeobox* (*WOX*) genes (Haecker, Groß-Hardt et al. 2004). In *A. thaliana* the *WOX* family comprises 15 members including *WUS* and *WOX1* to *14*, which

phylogenetic analysis grouped into three clades: the WOX13-clade composed of *WOX13*, *10* and *14*; the WOX9-clade, which includes *WOX8* and *9* as well as *11* and *12*; the WUS-clade consisting of the name giving *WUS* gene and *WOX1* to *7* (Haecker, Groß-Hardt et al. 2004; Deveaux, Toffano-Nioche et al. 2008; Nardmann and Werr 2012).

The clades originated at different time points during the evolution of land plants. Phylogenetic analysis places the WOX13-clade at the basis of the WOX family and all *WOX* genes detected in unicellular green algae and basal terrestrial plants such as the moss *Physcomitrella patens* relate to *WOX13* (Deveaux, Toffano-Nioche et al. 2008; Nardmann and Werr 2012). So far homologues of the other two WOX13-clade genes, *WOX10* and *14*, could only be identified in few Brassicacea species, an eudicot family, indicating them to be the product of rather recent gene duplication events. Thus the WOX13-clade, in particular *WOX13* itself, represent the ancestral lineage from which all other *WOX* genes evolved. Members of the other two non-ancestral WOX lineages, the WUS-clade and the WOX9-clade, can first be detected in basal vascular plants and at least one representative of each clade is apparent in leptosproangiate ferns that diverged prior to the emergence of seed plants (Nardmann and Werr 2012). In spermatophytes the family of *WOX* genes underwent a sudden and massive increase in number. Relatives of *WOX2*, *WOX3* and *WOX4* were identified in angiosperms and gymnosperms indicating their emergence at the basis of seed plants (Nardmann, Reisewitz et al. 2009). Also the two closely related genes *WUS* and *WOX5* appear to have evolved prior to the split of angiosperms and gymnosperms (Zhang, Jiao et al. 2017)

Aside of the aforementioned *WOX10* and *14*, several other genes seem to have developed rather recently during dicot plant evolution. So far homologues of *WOX8* and *WOX7*, which evolved within the *WOX9* and *WOX5* gene groups, respectively, as well as the two paralogous genes *WOX1* and *WOX6*, could be identified in Brassicacea, but apparently are absent in other families (Nardmann and Werr 2006; Vandenbussche, Horstman et al. 2009; Nardmann and Werr 2012).

Taken together the whole *WOX* gene family is likely to have emerged from a single gene similar to *WOX13* homologues of the ancestral lineage, which is also present in unicellular green algae. Group members of the WUS- and the WOX9-clade were isolated from leptosporangiate fern species. Thus at least a single member of each clade was present prior to the emergence of seed plants. Subsequent multiple duplication events in particular of genes belonging to the WUS-clade followed by neo- or subfunctionalization formed the diverse set of *WOX* genes in higher plants, which comprise e.g. 14 *WOX* genes in *A. thaliana*.

1.2.3 Functional divergence of WOX proteins

At least a single *WOX13* gene copy has been retained in all plants suggesting a conserved function of members of the ancient WOX lineage. The expression patterns and mutant analyses of *WOX13* homologues in unicellular algae, basal land plants and *A. thaliana* imply that *WOX13* may be associated with cell growth and division (Deveaux, Toffano-Nioche et al. 2008; Nardmann and Werr 2012; Sakakibara, Reisewitz et al. 2014). In the course of evolution, and thus the diversification of *WOX* genes, the comparable ubiquitous distribution of *WOX13* transcripts in basal plants appears to have become restricted to the domains of lateral root outgrowth and early flower development in *A. thaliana* (Deveaux, Toffano-Nioche et al. 2008). Gene expression of some members of the WOX9- and WUS-clade is confined to pluripotent cells during development of primary and lateral roots in the fern *Ceratopteris richardii* (Nardmann and Werr 2012). In contrast to their role in fern development, related WOX proteins in *A. thaliana* are required for proper embryonic development. Two WOX9-clade members, WOX9 and 8, function together with WOX2 of the WUS-clade in establishing the apical basal body axis of the *A. thaliana* embryo (Haecker, Groß-Hardt et al. 2004; Breuninger, Rikirsch et al. 2008). Other members of the WUS-clade are known primarily for controlling different aspects of meristematic activity during development of specialized structures and organs. In angiosperms expression of *WOX3* and/or *WOX1* homologues are required for the lateral outgrowth of leaves and floral organs (Matsumoto and Okada 2001; Nardmann, Ji et al. 2004; Shimizu, Ji et al. 2009; Vandenbussche, Horstman et al. 2009; Tadege, Lin et al. 2011; Nakata, Matsumoto et al. 2012). *WOX4* transcripts were detected in the procambium of *A. thaliana* and tomato, which harbours initial cells of the developing vasculature, and were shown to be necessary for the integrity and correct development of the vasculature (Hirakawa, Kondo et al. 2010; Ji, Strable et al. 2010). In the derived angiosperm species *A. thaliana* *WOX5* expression in the organizing centre of the root, the quiescent centre (QC), prevents premature differentiation of pluripotent stem cells in the root apical meristem (RAM) (Sarkar, Luijten et al. 2007). WUS takes on a similar role in the SAM and FM by maintaining a pluripotent stem cell pool (Mayer, Schoof et al. 1998) but is also required during ovule development of *A. thaliana* (Groß-Hardt, Lenhard et al. 2002). Gene expression of *WUS* and its homologues in the shoot apex differs partially between monocots and dicots, which may relate to divergences in SAM organization and thus in developmental programs (Nardmann and Werr 2006). Expression patterns of WUS/*WOX5* in gymnosperms indicate a conserved function during early stages of

sporogenesis and possibly also in propagating stem cell fate (Nardmann, Reisewitz et al. 2009). Even more so as the stem cell maintenance activity of WUS and WOX5 seems to have evolved prior to the split of fern and seed plants indicated by the results of interspecies complementation assays of the *wus-1* and *wox5-1* loss of function mutations in *A. thaliana* (Zhang, Jiao et al. 2017).

In the light of this diverse set of roles of WOX proteins during plant development the question arises about the evolution and the molecular basis of diverging functions. The two paralogous genes *WOX8* and *9* act redundantly during embryonic development of *A. thaliana* (Breuninger, Rikirsch et al. 2008). Similarly most of the WUS-clade proteins of seed plants were shown to be able to functionally substitute for each other, some were even tested successfully across species (Sarkar, Luijten et al. 2007; Shimizu, Ji et al. 2009; Lin, Niu et al. 2013; Dolzblasz, Nardmann et al. 2016; Zhang, Jiao et al. 2017). The main requirement in these cases is transcriptional control by the appropriate promoter indicating that the evolution of divergent expression patterns may be the predominant cause for the differing functions of WUS-clade proteins in seed plants. However, the ability to functionally substitute for each other does not seem to extend beyond clade boundaries. Neither is WUS able to replace WOX9 and 8 during embryogenesis nor are WOX proteins outside the WUS-clade able to substitute for WOX1 in tobacco or WUS in *A. thaliana* (Lin, Niu et al. 2013; Dolzblasz, Nardmann et al. 2016). Partially these functional differences seem to relate to the acquisition of repressive transcriptional activity conferred by the WUS-box (Lin, Niu et al. 2013). The WUS-box is apparent only in members of the WUS-clade (reviewed in van der Graaff, Laux et al. 2009) and was shown to be essential for interaction with the transcriptional co-repressor TOPLESS (TPL), which is also called WUSCHEL INTERACTING PROTEINS 1 (WISP1) (Kieffer, Stern et al. 2006; Szemenyei, Hannon et al. 2008). Interestingly fusion with either the WUS-box or SRDX, an artificial repressive motif (Hiratsu, Matsui et al. 2003), is sufficient to transfer the ability to substitute WOX1 function in tobacco to WOX9-clade proteins (Lin, Niu et al. 2013). However, the same fusion with WOX13 only allows partial complementation of *wox1* mutant tobacco plants. Furthermore WUS function in SAM and FM of *A. thaliana* can not be rescued by protein chimeras consisting of the WUS-box fused to WOX13 or members of the WOX9-clade (Dolzblasz, Nardmann et al. 2016). Homeodomain swaps in context of full length WUS protein revealed a partial capability of the WOX8 and WOX9 homeodomains to replace the WUS homeodomain, albeit with apparent lowered transcriptional activity, whilst the

WUS-WOX13 HD chimera fails to rescue any aspect of the *wus* loss of function mutation phenotype in *A. thaliana* (Dolzblasz, Nardmann et al. 2016).

These results indicate that some differences in WOX protein function relate to alterations of expression patterns or in the divergent recruitment and interaction of co-factors. Others, however, can be directly linked to the WOX homeodomain. Thus evolution of functional divergence in the WOX family, in particular between the clades, seems to be associated with alterations of the homeodomain. Furthermore the molecular changes differentiating WOX13- and WUS-clade homeodomains seem to exert greater effects than those of the homeodomains of the two non-ancestral clades.

2 Materials and methods

2.1 Chemicals

chemical	abbreviation	manufacturer
agar		Roth
agarose		Roth
ammonium persulfate	APS	Sigma
boric acid		Roth
bovine serum albumin	BSA	Sigma
bromophenol blue	BPB	Sigma
dithiothreitol	DTT	Applichem/Sigma Aldrich
ethidium bromide		Sigma Aldrich
ethylenediaminetetraacetic acid	EDTA	BDH Prolabo
glacial acetic acid		Roth
glucose		Roth
glycerol		Sigma Aldrich
glycin		Sigma
imidazole		Fluka
isopropyl β -D-1-thiogalactopyranoside	IPTG	Duscheffa Biochemica
magnesium chloride	MgCl ₂	Merck
<i>N-N'-N'</i> tetramethylethan-1,2-diamin	TEMED	Sigma Aldrich
potassium chloride	KCl	Sigma Aldrich
Rotigel 30 (37, 5:1)		Roth
sodium chloride	NaCl	Roth
sodium dodecyl sulfate	SDS	Roth
sodium hydroxide	NaOH	Merck
tris		Roth
tryptone		Merck

2.2 Vectors

Cloning	pJET1.2/blunt	Thermo Scientific
Protein expression	pET22b	
	pStrep-1	
	pET28a	
	pET28aN-StrepII-Thrombin	
	pTYB2	NEB
	pTXB1	NEB

2.3 Kits

Plasmid DNA purification	Machery-Nagel
PCR clean-up & Gel extraction	Machery-Nagel
BCA Protein Assay Kit	Pierce

2.4 Oligonucleotides

Primers for cloning:

name	sequence 5' to 3'
pTYB2_SmaI_Fw	CCCGGGTGCTTTGCCAAG
pTYB2_NdeIXhoI_Rv	CCGCTCGAGCATATGTATATCTCCTC
T7	TAATACGACTCACTATAGGG
pTXB1_Rv	CTTTCAGGTCGATGG
pTYB2_Rv	TTAATTACCTCACGAGGT
seqpET_F	CACGATGCGTCCGGCGTAGAGG
seqpET_R	CCGTTTAGAGGCCCAAGGGGTTATG
AtWusFLNdeI_F2	GGAATTCCATATGGAGCCGCCACAGCA
AtWusFLXhoI_stop_R2	CCGCTCGAGCTAGTTCAGACGTAGCTCAA
AtWusFLXhoI_nostop_R2	CCGCTCGAGGTTTCAGACGTAGCTCAA
AtWox13FLNdeI_F2	GGAATTCCATATGATGGAATGGGATAATCAGCTACAAC
AtWox13FLXhoI_stop_R2	CCGCTCGAGTCAGCCTGACATGCCATAAT
AtWox13FLXhoI_nostop_R2	CCGCTCGAGGCCTGACATGCCATAAT

Oligonucleotides for ITC:

name	sequence 5' to 3'	gene	organism	reference
ag_Fw	TGGATTTATACCCAATGTGTTAATGGGTTGT	<i>AGAMOUS</i> ; 2nd intron	<i>Arabidopsis thaliana</i>	(Lohmann, Hong et al. 2001)
ag_Rv	ACAACCCATTAACACATTGGGTATAAATCCA			
mag_Fw	TGGATTTATACCCAATGTGTTCTGGGTTGT			
mag_Rv	ACAACCCAGGAACACATTGGGTATAAATCCA			
agwobble_Fw	TGTTAAACTTCGACGTATGTAGTTGTGTGTA			
agwobble_Rv	TACACACAACCTACATACGTCGAAGTTTAACA			
clv3-1080_Fw	GGCTCATATAATCCATTCAATTTATG	<i>CLAVATA 3</i> ; -1080	<i>Arabidopsis thaliana</i>	(Yadav, Perales et al. 2011)
clv3-1080_Rv	CATAAATTGAATGGATTATATGAGCC			
mclv3-1080_Fw	GGCTCATATAGGCCATTCAATTTAG			
mclv3-1080_Rv	CTAAATTGAATGGCCTATATGAGCC			
g-boxnest_Fw	CTGGACAATTCACGTGAGCAGTCACT			(Busch, Miotk et al. 2010)
g-boxnest_Rv	AGTGACTGCTCACGTGAATTGTCCAG			

2.5 Bacteria strains

strain	genotype	reference
DH5 α	F- <i>endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20 ϕ80dlacZΔM15 Δ(lacZYA-argF)U169, hsdR17($r_K^-m_K^+$), λ-</i>	
BL21 (DE3)	F- <i>ompT gal dcm lon hsdS_B($r_B^-m_B^-$) λ(DE3 [<i>lacI lacUV5-T7 gene 1 ind1 sam7 nin5</i>])</i>	
BL21 star (DE3)	F- <i>ompT gal dcm lon hsdS_B($r_B^-m_B^-$) rne131 λ(DE3 [<i>lacI lacUV5-T7 gene 1 ind1 sam7 nin5</i>])</i>	
ER2566	F-, λ -, <i>fhu A2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mc-73::miniTn10--Tet^S)2 [dcm] R(zgb-210::Tn10--Tet^S) endA1 Δ(mcrC-mrr)114::IS10</i>	
M15 [pREP4]	F-, Φ 80 Δ lacM15, thi, lac-, mtl-, recA ⁺ , KmR	Valerjo & Zabin in 1974 (Villarejo and Zabin 1974)

2.6 Media for bacterial cell cultures

All media were autoclaved at 181 °C for 15 minutes. For agar plates 1.5% (w/v) agar was added to the medium prior to the autoclaving process.

LB (lysogeny broth) medium 1L:	10 g Tryptone
	5 g Yeast extract
	10 g NaCl
SOC medium 1L:	20 g Tryptone
	5 g Yeast extract
	0.5 g NaCl
	2.5 mM KCl
	20 mM Glucose
	10 mM MgCl ₂

Glucose and MgCl₂ were added from sterile filtered stock solutions after all other ingredients had been autoclaved.

2.7 Buffers

5x TBE	450 mM Tris 450 mM Boric Acid 10 mM EDTA pH 8.0
50x TAE	2 M Tris 5.71 % (v/v) glacial Acetic Acid 50 mM EDTA pH 8.0
SDS-PAGE 3x SDS-sample buffer	180 mM Tris-HCl pH 6.8 19.2 % (v/v) Glycerol 6 % (w/v) SDS 0.3 M DTT 0.075 % (w/v) BPB
4x separation gel buffer	1.5 M Tris-HCl pH 8.8 0.4 % (w/v) SDS
4x stacking gel buffer	0.5 M Tris-HCl pH 6.8 0.4 % (w/v) SDS
10x running buffer	250 mM Tris 1.92 M Glycine 0.5 % (w/v) SDS
<i>Purification of WOX13 fused to His tag</i>	
pellet wash buffer	20 mM Tris-HCl (pH 7.5) 150 mM NaCl
lysis buffer	20 mM Tris-HCl (pH 7.5) 200 mM NaCl 10 mM imidazole (pH 7.5)
wash/elution buffer	20 mM Tris-HCl (pH 7.5) 200 mM NaCl and varying concentrations of imidazole (pH7.5) wash20: 20 mM wash50: 50 mM wash100: 100 mM wash250: 250 mM

SEC buffer	composition identical to lysis buffer filtered and degased
------------	---

Purification of WOX13 fused to MxeInt/CBD tag

MxeInt: mini-intein translated from the modified *gyrA* gene of *Mycobacterium xenopi*

CBD: chitin binding domain

MxeInt/CBD: translational fusion tag

pellet wash buffer	20 mM Tris-HCl (pH 8.5) 300 mM NaCl
lysis buffer	20 mM Tris-HCl (pH 7.5) 300 mM NaCl
elution buffer	identical to lysis buffer
SEC buffer	20 mM Tris-HCl (pH 7.5) 150 mM NaCl filtered and degased

Purification of Wuschel HD and WOX13 HD fused to SceInt/CBD tag for crystallisation

SceInt: intein translated from VMA1 gene of *Saccharomyces cerevisiae*

CBD: chitin binding domain

SceInt/CBD: translational fusion tag

pellet wash buffer	20 mM Tris-HCl (pH 8.5) 300 mM NaCl
lysis buffer	20 mM Tris-HCl (pH 8.5) 500 mM NaCl
elution buffer	identical to lysis buffer
SEC buffer	20 mM Tris-HCl (pH 8.5) 300 mM NaCl 1 mM DTT filtered and degased

Purification of Wuschel HD and WOX13 HD fused to SceInt/CBD tag for ITC

pellet wash buffer	20 mM Tris-HCl (pH 8.5) 300 mM NaCl
lysis buffer	20 mM Tris-HCl (pH 8.5) 500 mM NaCl

Transfer buffers (TF)	mixed from lysis buffer and SEC buffer	
	parts lysis buffer	parts SEC buffer
TF I	4	1
TF II	3	2
TF III	2	3
TF IV	1	4
elution buffer	TF IV (see table above)	
SEC buffer	20 mM Tris-HCl (pH 8.5)	
	100 mM NaCl	
	filtered and degased	
ITC buffer	identical to SEC buffer	

2.8 Amino acid frequency maps

Amino acid sequences were aligned with ClustalW Multiple Alignment in BioEdit at default settings (Thompson, Higgins et al. 1994) and frequency maps were generated with weblogo (Schneider and Stephens 1990; Crooks, Hon et al. 2004).

WOX homeodomains comprise 41 WUS, 55 WOX3, 53 WOX8 and WOX9 as well as 81 WOX13 sequences from moss, fern, angiosperms and gymnosperms, which were extracted and sorted in a phylogenetic tree by Viera Kovacov (Department of Cellular networks and systems biology at the university of Cologne).

The bit score map of animal homeodomains is consists of 1628 sequences from Nematods, Arthropods, the basal vertebrate Amphioxus as well as representatives of fish, amphibians, birds and mammals, sequences were obtained from the Homeobox Database HomeoDB2 (Zhong, Butts et al. 2008; Zhong and Holland 2011) and include homeodomain sequences from the ANTP megacluster (HOX, ParaHOX and NK) and the classes PRD, LIM, POU, HNF, SINE, TALE, CUT and PROS, sequences were aligned to Drosophila HOX homeodomains and unaligned residues were cut to fit the typical 60 amino acid homeodomain sequence.

The sequence logo of plant homeodomains is modified after a compilation of plant homeodomains published by Mukherjee and colleagues (Fig. 1 in Mukherjee, Brocchieri et al. 2009). WOX homeodomain sequences were excluded and apparently conserved glycine residues between helices II and III were aligned prior to the generation of the frequency map.

2.9 Gelelectrophoresis

2.9.1 Agarose gel electrophoresis

Appropriate amounts of agarose were weighted in to obtain the desired concentration and resolved in 1x TAE. For the purpose of visualizing the DNA later on 0.5 mg/ml EtBr were added. Gels were prepared and run utilizing the PerfectBlue Gel System Mini M (peqlab).

2.9.2 Polyacrylamide gel electrophoresis (PAGE)

A 30% acrylamide solution (Rotigel 30; 37.5:1; Roth) was diluted in the appropriate buffer to the desired concentration. The polymerization process was started by the addition of 0.1 % (v/v) *N-N-N'-N'* tetramethylethan-1,2-diamin (TEMED) and 0.15 % (w/v) ammonium persulfate (APS). The gels were prepared and run utilizing the systems for mini-gels supplied by BioRad.

In the case of native PAGE the gels were prepared and run in pre-cooled 0.5x TBE. Prior to the addition of samples the gels were pre-run at 100 V and 4 °C for 20 minutes. Subsequently the samples mixed with DNA-loading buffer were applied, run at 100 V for another 15 minutes after which the voltage was increased to 120 V.

Denaturing gels were prepared as discontinuous system, with approximately 1 cm stacking gel (4.2 % acrylamide in 1x stacking buffer) above the separation gel prepared in 1x separation buffer. Samples were mixed with SDS-sample buffer and boiled at 95 °C for 4-5 minutes, applied to the gel and run in running buffer at 25 mA.

To detect DNA gels were stained with 0.5 mg/ml EtBr in 0.5x TBE for 5-10 minutes, washed 3 times in 0,5x TBE and illuminated with UV light (Gel Doc XR+, Biorad).

Proteins were visualized with Coomassie Brilliant Blue (Rotiblu, Roth) staining according to manufacturer's instructions. Images were recorded with trans-UV and trans-white conversion screen (Gel Doc XR+, Biorad).

2.10 Cloning

The genes *WUSHD* and *WOX13HD*, which translate into 66 and 65 aminoacids, respectively, were subcloned into the vector pTYB2 (NEB) via the restriction sites for *SmaI* and *NdeI* by Emmi Wachsmut. Those constructs will be referred to as *WUSHD*-pTYB2 and *WOX13HD*-pTYB2 hereafter.

Primers were designed to amplify the vector pTYB2 of the construct *WUSHD*-pTYB2 and simultaneously introduce restriction sites for the enzymes *NdeI*, *XhoI* and *SmaI*. The PCR products were purified with agarose gel electrophoresis and the ends ligated with T4 DNA-ligase (Invitrogen). DH5 α *E. coli* cells were transformed with the ligation mix and positive transformants were selected on LB-agar plates supplemented with 100 μ g/ml ampicillin. Colony PCR was performed to check for clones containing the desired plasmid. Positive clones were grown in LB-medium and 100 μ g/ml ampicillin at 37 °C over night. Plasmids were purified to be used in further cloning processes and are labelled as pTYB2.02 from in the following.

To obtain constructs for the expression of tagged full-length protein, plasmids containing the complete coding sequence of either *WUSCHEL* or *WOX13* were used as template for amplification. The primers used in this amplification process were designed to simultaneously introduce *NdeI* and *XhoI* restriction sites at the N-terminus and the C-terminus, respectively, as well as a stop codon if the tag was to be fused to the N-terminus. Purified PCR-products were ligated with pJET1.2/blunt by T4-DNA ligase (Invitrogen). The ligation mix was used to transform DH5 α *E. coli* cells. Positive clones were identified with colony-PCR and grown in LB-medium containing 100 μ g/ml ampicillin over night. Plasmids were purified and sequenced at GATC-Biotech. Plasmids containing the correct insert were digested with *NdeI* and *XhoI* (both NEB). The insert was separated by agarose gel electrophoresis and extracted. The purified insert was ligated with an expression plasmid processed the same way as the insert overnight at 16 °C. The ligation mix was used to transform DH5 α *E. coli* cells. Clones containing the insert were identified with colony-PCR and grown in LB-medium containing the appropriate antibiotics over night. Plasmids were prepared and sequenced at GATC-Biotech.

2.11 Transformation

To transform bacterial cells 100-200 ng plasmid DNA were mixed with 30 μ l chemically competent cells and incubated on ice for 10-15 minutes. Subsequently the cell-plasmid mixture was incubated in a water bath at 42 °C for 30 seconds. After the addition of 270 μ l SOC medium the cells were allowed to recover in a rotating wheel at 37 °C for 1 hour. Positive clones were selected on LB-Agar plates supplemented with the appropriate antibiotics. Glucose was added to a final concentration of 20 mM, only if the transformed cells were to be used for protein expression in SOC medium later on.

2.12 Test expressions

Testexpressions were conducted with 40 ml SOC- or LB-medium supplemented with the appropriate antibiotics. The medium was inoculated with freshly transformed cells and incubated at 37 °C. When the cell suspension had reached an OD₆₀₀ of 0.6-0.8 recombinant protein expression was induced by the addition of 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein expression was analysed at 37 °C, 28 °C and 16 °C for 3, 5 and 16 hours, respectively. All samples were normalized to the OD₆₀₀ of the corresponding 1 ml sample taken prior to induction. Cells were centrifuged at 14000 xg and 4 °C for 5 minutes. The pellets were resuspended in 70 µl lysis buffer. A 50 µl aliquot was sonicated in a resonance cup filled with 4 °C cooling agent (Branson Sonifier). Half of the lysate was centrifuged at 14000 xg and 4 °C for 5 minutes, the supernatant was transferred to a fresh eppendorf tube and the pellet resuspended in 25 µl lysis buffer. Samples were analysed with SDS-PAGE.

2.13 Large scale protein expression

SOC- or LB-medium supplemented with the appropriate antibiotics was inoculated with single colonies of freshly transformed cells. Bacteria were grown at 28 °C or 37 °C over night. These starter cultures were diluted into fresh medium to an OD₆₀₀ of 0.1 and grown at 37 °C until an OD₆₀₀ of 0.6-0.8. In case the protein expression was conducted at lower temperatures, the incubator was cooled down prior to induction. Protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. After protein expression had continued for the time indicated below the cells were harvested at 4000 rpm and 4 °C for 20 minutes (Sorvall RC12 BP; ThermoScientific). The supernatant was discarded, pellets were resuspended in pellet wash buffer, transferred into Falcon tubes and again centrifuged at 4 °C and 4000 rpm for 20 minutes (Heraeus Multifuge X1R; ThermoScientific). Supernatant was removed and the pellets were stored at -80 °C until usage. To monitor the protein expression samples of the cell suspension were taken prior to the addition of IPTG and just before cell harvest and analysed with SDS-PAGE.

2.14 Protein purification

All following steps were conducted at 4 °C and only pre-cooled buffers were used.

2.14.1 Cell lysis and clarification of lysate

Pellets were thawed on ice and resuspended in pre-cooled lysis buffer supplemented with DNaseI to a final concentration of 10 µg/ml. The cell suspension was processed twice in a cell disruptor (TS-Series Cell Disruptor, Constant Systems Ltd.) at 2.5 kbar with precooled pressure head. To remove the cell debris the lysate was centrifuged at 40000 rpm and 4 °C for 40 minutes (XL-70 or Optima TM L-80 XP Ultracentrifuge, both Beckmann Coulter). The supernatant was used for the next purification steps.

2.14.2 Affinity purification via Int/CBD tag

The tag fused to the C-terminus of the protein consists of two parts. As the name implies, the chitin binding domain (CBD) is required for the purification of the target protein by binding to chitin beads (IMPACT TM; NEB). The intein, which links the target protein and the CBD, allows the elution of the target protein from the column. This is achieved by the addition of a thiol reagent, which induces the intein's self cleavage activity.

The soluble part of the cell lysate was applied to a column (inner diameter: 1 cm) filled with chitin beads (IMPACT TM; NEB) using a peristaltic pump at a speed between 0.6-0.9 ml/min. The column was washed until the absorption at 280 nm of the flow through reached the base line and flushed with at least 3 bed volumes elution buffer supplemented with 50 mM DTT to guarantee an even distribution of the thiol reagent. The flow was stopped and the column incubated at 16 °C over night. The next day cleaved target protein without tag was eluted in elution buffer. The UV absorption of the flow through was monitored to determine the end of the elution process.

2.14.3 Affinity purification via 6xHis tag

The supernatant was applied to NiNTA-beads (Superflow; Quiagen) equilibrated in lysis buffer using a peristaltic pump and washed with lysis buffer until the UV-absorption of the flow through had returned to baseline levels. Subsequently the column was flushed with lysis buffer supplemented with increasing imidazole concentrations. Aliquots were analysed on SDS-PAGE. The protein eluted with 250 mM imidazole consisted of a single band at the correct size and was used in further purification steps.

2.14.4 Size exclusion chromatography

The protein eluted from the chitin or Ni-NTA column was concentrated to 1.8-2.2 ml in a centrifugal filter unit (Vivaspin 20, 3000 MWCO PES; Sartorius) at 3500-4000 xg and 4 °C in intervals of 30-60 minutes (Heraeus Multifuge X1R; ThermoScientific). Size exclusion chromatography was performed on HiLoad™ 16/600 Superdex 200 prep grade and HiLoad™ 16/600 Superdex 75 prep grade (GE Healthcare) columns in case of full length protein and the homeodomain, respectively. The columns were run with ÄKTA systems (ÄKTAprime, ÄKTAprime plus, ÄKTA purifier; GE Healthcare) at 4 °C and a speed of 1 ml/min. Fractions of 1 ml were collected starting at a volume of 40 ml until 120 ml. The chromatogram was analysed with UNICORN software (GE Healthcare) and fractions were analysed on SDS-PAGE.

2.15 Protein crystallisation

Using the chromatogram fractions containing clean and supposedly uniform protein were identified. Those fractions were fused and concentrated in a centrifugal filter unit (Vivaspin 6, 5000 MWCO PES; Sartorius) at 3500-4000 xg and 4 °C in intervals of 30-60 minutes (Heraeus Multifuge X1R; ThermoScientific) to a final concentration of 9-17 mg/ml. Concentrations were determined via UV-absorption at 280 nm and the respective extinction coefficients calculated in ProtParam (<http://web.expasy.org/protparam>). Screens from different manufacturers were used to search for crystallisation conditions. All crystals were grown by the sitting drop vapor diffusion technique.

Table 3: Conditions under which WUSHD crystals formed

salt	buffer	protein concentration	T	duration
1.2 M sodium citrate tribasic dehydrate	0.1 M BTP pH 7	16.1 mg/ml 13.7 mg/ml	4 °C 20 °C	2-3 months
1.2 M sodium citrate tribasic dehydrate	0.1 M Tris pH 8.5	16.1 mg/ml 13.7 mg/ml	4 °C 20 °C	
1.2 M D,L-Malic acid pH 7	0.1 M BTP pH 7	9 mg/ml	20 °C	2 weeks, when seeded
1.8 M Sodium phosphate monobasic monohydrate, Potassium phosphate dibasic pH 8.2		16.1 mg/ml	20 °C	half a year, very small
2.1 M DL-Malic acid pH 7		9 mg/ml	20 °C	3 months, very small
				half a year, very small

WUSCHEL homeodomain was mixed with precipitant in a ratio 1:1 to form 200 nl drops and incubated at 20 °C or 4 °C. After a few days needle clusters formed in 0.1 M Bis-Tris-propane (BTP) pH 7 with 60 % Tacsimate pH 7 with high consistency at different

protein concentrations (9, 13.7 and 16.1 mg/ml) at 20 °C as well as at 4 °C. Other conditions for the formation of WUSCHEL homeodomain crystals are listed in Table 3.

2.16 Optimization of WUSCHEL homeodomain needle cluster

A grid screen composed of 0.1 M Bis-Tris-propane (BTP) pH 6.8-7.4 and 52.5-65 % Tacsimate pH 6.8-7.4 was prepared and protein (17.2 mg/ml) was mixed with precipitant in a ratio 1:1 or 2:1 to form 2 or 3 µl drops, respectively. Seeds were prepared from a needle cluster by rigid mixing (vortex for 1 minute) in 40 µl and diluted 1:1000 in the original reservoir solution. To yield well-ordered needle shaped crystals in several conditions 0.5 µl of seeds were added to each well of the grid screen. Native data were collected from a crystal grown in 0.1 M BTP pH 7 and 52.5 % Tacsimate (pH 7) that was cryo-protected with 0.1 M BTP (pH 7), 52.5 % Tacsimate (pH 7) and 25 % sucrose. Native crystals grown in the same condition were soaked with 10 mM K₂O₄Os for 30 minutes to derivatize the crystals for phasing.

2.17 Data collection and determination of the WUSCHEL homeodomain structure

Native data were collected at the European Synchrotron Radiation Facility, Grenoble, France, on beamline ID23.2. Data from heavy atom derivatized crystals were collected at the Swiss Light Source, Paul-Scherrer-Institute, Villigen, Switzerland, on beamline X06DA. A peak dataset at a wavelength of 1.1399 Å and a native dataset at 1.20 Å were collected.

All diffraction data were processed using XDS (Kabsch 2010). The structure was solved by SAD using phases computed from the three osmium sites by the phenix.autosol routine of the PHENIX package (Adams, Afonine et al. 2010) and an almost complete model was automatically built. The model was refined using iterative cycles of phenix.refine (Afonine, Grosse-Kunstleve et al. 2012) and manual model building in Coot (Emsley, Lohkamp et al. 2010). The structure was solved by molecular replacement with one monomer from the first model as search model by the program Phaser (McCoy, Grosse-Kunstleve et al. 2007) and refined as described above. All data collection and refinement statistics are given in Table 4.

Table 4: data collection and refinement statistics; values in parentheses are for the highest resolution shell

Crystal parameters	Native (ESRF)	Native	Os-edge (anom)
Space group	$P2_12_12$	$P2_12_12$	$P2_12_12$
Unit-cell parameters (Å, °)	a=62.00, b=90.12, c=23.25, $\alpha=\beta=\gamma=90.00$	a=61.99, b=89.97, c=23.23, $\alpha=\beta=\gamma=90.00$	a=62.85, b=90.49, c=23.25, $\alpha=\beta=\gamma=90.00$
Z (molecules per ASU)	1	1	1
Matthews coefficient (Å ³ Da ⁻¹)	1.92	1.90	1.90
Solvent content (%)	35.2	35.29	35.29
Data-processing statistics			
Temperature of measurement (K)	100	100	100
Wavelength (Å)	0.8726	1.20	1.1399
Resolution (Å)	45.18-1.85 (1.98-1.85)	45.18-2.19 (2.32-2.19)	45.25-2.48 (2.63-2.48)
Total reflections	39334 (6827)	67537 (10475)	62934 (9397)
Unique reflections	11657 (2049)	7328 (1154)	9131 (1459)
Completeness (%)	98.8 (96.8)	99.6 (97.8)	99.6 (97.6)
R _{merge} (%)	11.8 (86.6)	17.6 (67.7)	
Mean I/σ(I)	9.83 (1.56)	12.57 (3.67)	9.39 (2.71)
Refinement statistics			
Resolution range (Å)	36.41-1.85		
R _{work} /R _{free} † (%)	18.15/22.47		
No. of non-H protein atoms	1015		
No. of water molecules	113		
Root-mean-square deviations			
Bond lengths (Å)	1.236		
Bond angles (°)	0.014		
Average B factor (Å ²)			
All protein atoms	26.35		
Waters	33.06		
Ramachandran plot (%)			
Most favored	95.65		
Additional allowed	4.35		
Generously allowed	0.0		
Generously allowed	0.0		

†random 5% of the working set of reflections

2.18 Homology modelling of other WOX homeodomains

Models of other WOXHDs were built in Modeller 9.13 (Sali and Blundell 1993) using the graphical interface easy modeller 4.0 (Kuntal, Aparoy et al. 2010) and automatic loop refinement. Chain A of the WUSHD structure served as template. Modelled structures were verified by their Ramachandran plot. In case of WOX4, 5 and 8 single outliers were manually refined in Coot (Emsley, Lohkamp et al. 2010).

2.19 Structural analysis

The structures were analysed with Coot (Emsley, Lohkamp et al. 2010) and ccp4mg version 2.10.4 (McNicholas, Potterton et al. 2011). The accessible surface area (ASA) was calculated in ccp4mg (McNicholas, Potterton et al. 2011) with a water radius of 1.4 Å and point density set to 1 Å. To obtain relative ASAs the values of each residue or the side chain were normalized to the respective maximum surfaces as determined for Gly-X-Gly tripeptides by Miller et al. (Miller, Janin et al. 1987). The webtool RAPIDO (Mosca, Brannetti et al. 2008; Mosca and Schneider 2008) was used for pairwise alignment of selected homeodomains (Table 5) to chain A of WUS HD. In all cases the LoLim factor was set to 1.5, except for the alignment of HNF1β HD to WUS HD, where it was set to 1. Images were taken with PyMol Molecular Graphics System (Version 1.3; Schrödinger LLC).

Table 5: Homeodomains used for structural comparison to WUS HD

Homeodomain of	Abbreviation	PDB ID	Chains & Residues extracted from PDB file	Used to determine	
				diverging regions	ASA
ENGRAILED	EN HD	1ENH 3HDD	chain B	x	x x
MATING TYPE α1	MATα1 HD	1YRN 1AKH	chain A chain B	x	x x
EXTRADENTICLE	EXD HD	1B8I	chain B	x	x
MATING TYPE α2	MATα2 HD	1APL 1YRN	chain C chain B	x	x x
HEPATOCYTE NUCLEAR FACTOR β1	HNF1β HD	2H8R	chain A, residues 231-308 chain B, residues 231-310	x	x x

3 Results

3.1 Description of the crystal structure of the WUSCHEL homeodomain

3.1.1 Overall structure

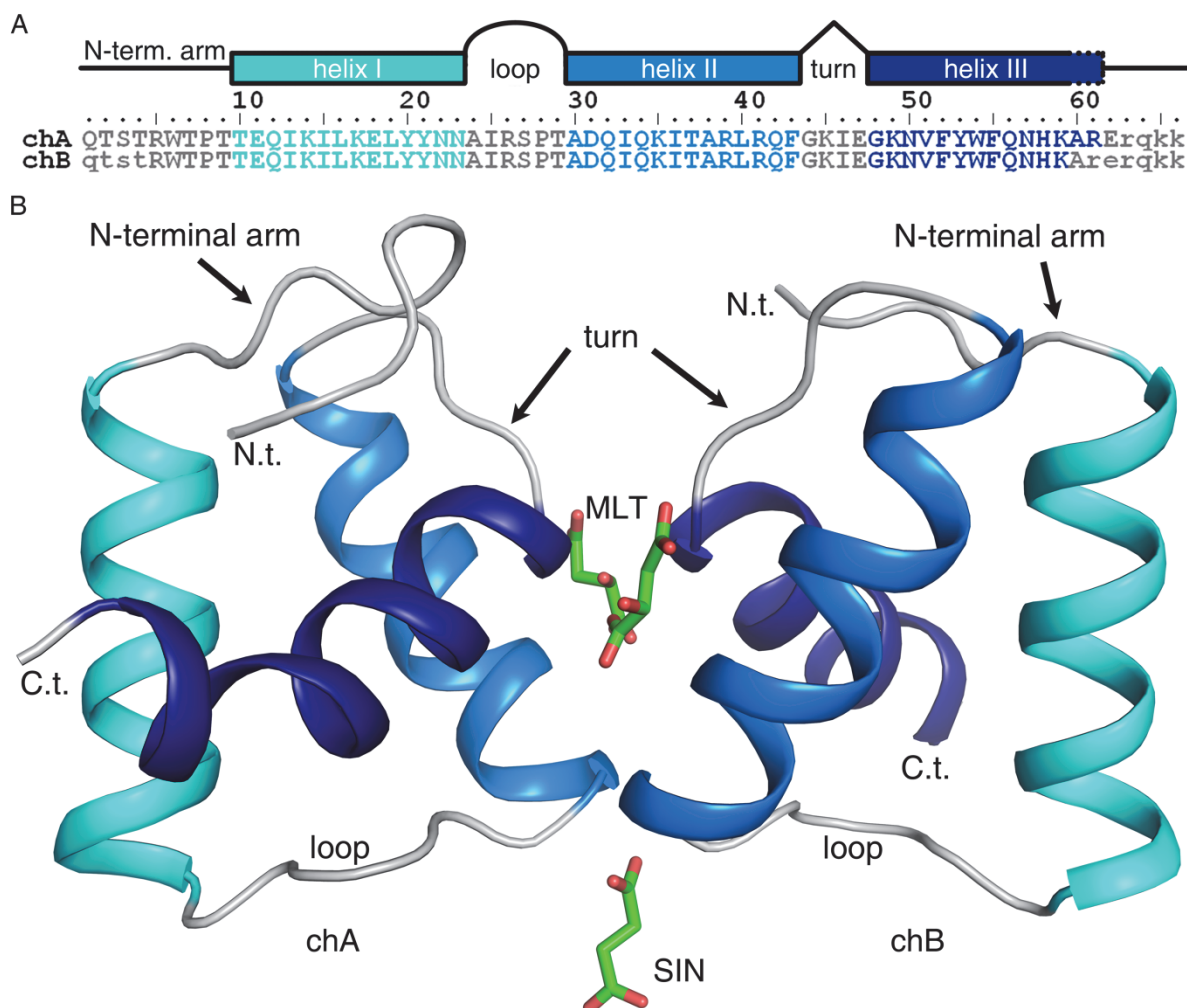


Fig. 4: Overall structure of WUS HD. A) Scheme of WUS HD; helices are indicated by boxes; resolved residues of chain A (chA) and chain B (chB) are written in capital letters, unresolved in small letters; N-terminal arm is abbreviated as “N-term. arm” B) The ASU of the WUS HD crystal; two WUS HD copies are packed in one ASU and were co-crystallized with two malate molecules (MLT) and one succinate molecule (SIN); C- and N-terminus are abbreviated as “C.t.” and “N.t.”, respectively.

The crystal of the WUSCHEL homeodomain (WUS HD) was diffracted to a resolution of 1.85 Å. The structure was solved using phases from three osmium sites determined by single wavelength anomalous dispersion and was refined to an R_{work} of 18.15 % and an R_{free} of 22.47 % (Table 4). Two independent WUS HD copies (chains A and B) are packed in one

asymmetric unit (ASU), which were co-crystallized with two malate (MLT) molecules and one succinate (SIN) molecule in the native crystal (Fig. 4B).

A model could be placed in the electron density map ranging from residues Gln1 up to Glu62 in chain A (chA) and Arg5 to Ala60 in chain B (chB), respectively (Fig. 4B). The disparate resolution of terminal residues between the two copies is likely to be caused by differences in the packing of chA and chB in the crystal. In chB both termini project into the solvent, whereas in chA they are stabilized by crystal contacts to neighbouring molecules. Still the general structural composition typical for homeodomains is well defined in both copies. The structure comprises a largely flexible N-terminal arm followed by three α -helices, which are arranged around a hydrophobic core (see section 3.1.4) to form a globular shaped protein. The most N-terminal helix, helix I (residues Thr10 to Tyr21), is separated from helix II (residues Ala30 to Phe43) by a loop. Helix I and II are arranged in an anti-parallel manner, whereas a turn of four amino acids places the third helix (helix III) roughly perpendicular to helix II. Helix III, also called the DNA recognition helix, starts at residue Gly48 and is resolved until residue Arg61 (chA) and Lys59 (chB), respectively.

3.1.2 Non-crystallographic symmetry

Two independent copies of WUS HD are packed in one asymmetric unit. Analysis of non-crystallographic symmetry (NCS) is possible for residues Arg5 to Ala60, which are resolved in both chains (Fig. 4A). The two copies were superposed and distances between C α -atoms of corresponding residues were measured in PyMOL. Additionally the mean distance (root mean square deviation, RMSD) of individual structural elements was determined without fitting. The two WUS HD copies in the crystal are nearly identical (Fig. 5A and C). This is also reflected in very low RMSD values even when the calculation is not limited to C α -atoms, but includes all atoms (Table 6). As expected the highest variation occurs at the N- and C-terminus of the homeodomain, where structures in general tend to be more flexible (Fig. 5A and C).

The recognition helix of chB follows the straight path of a α -helix, with the secondary amide of a peptide bond donating its hydrogen to the main chain's carbonyl group of the residue located four positions earlier (closer to the N-terminus) (Fig. 5B), whereas the recognition helix of chA displays a kink after His58. The usual interaction between the main chains of Lys59 and Phe55 is disrupted (distance: 5 Å). Instead the helix is stabilized by a network of polar interactions between the backbone carbonyl groups of Asn57 and His58 and NH-groups of Ala60 and Arg61 (Fig. 5B). Given the general high flexibility of the N-terminal arm, the obvious differences in

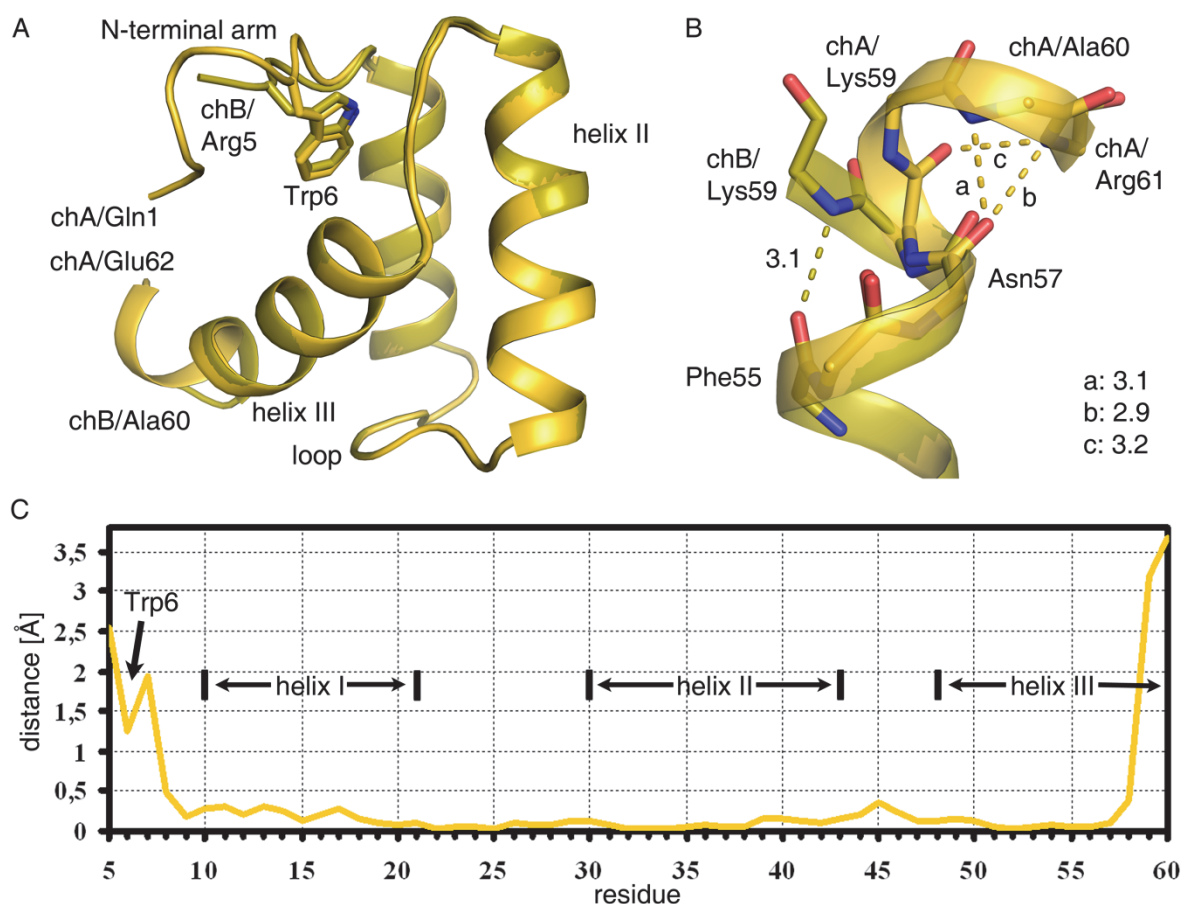


Fig. 5: Non-crystallographic symmetry of WUS HD. A) and B) Superposed structures of chA (yellow, residues Gln1 to Glu62) and chB (olive, residues Arg5 to Ala60) displayed as ribbons. B) close up on C-terminus of helix III; the main chains of residues Phe55 up to Arg61 (chA, yellow) or Lys59 (chB, olive), respectively, are depicted as sticks and ribbon; interactions stabilizing the C-terminus of helix III in chA and chB are represented as dashed lines; distances are given in [Å]. C) Distance of C α -atoms (y-axis) of corresponding residues Arg5 up to Ala60 (x-axis) in chA and chB.

conformation of this structure between the two WUS HD copies are not very surprising. However, the range of motion of the N-terminal arm seems to be partially constrained by Trp6 (Fig. 5A). At this residue the two chains converge shortly before increasing the distance again on either side of Trp6 (Table 6). Interestingly the relative positions of the two side chains of Trp6 are nearly identical though the locations of their C α -atoms deviate by 1.2 Å (Fig. 5A). This finding relates to contacts established by the Trp6 side chain in docking of the N-terminal arm, which will be discussed in more detail in section 3.1.5.

The loop of WUS HD is somewhat irregular in structure but contains two consecutive type IV β -turns, spanning residues 21-24 and 25-28, respectively, although the hydrogen bond between residues 25 and 28 is absent. In this regard the loop of WUS HD appears unusually rigid, as is reflected by the extremely low divergence in this region between chA and

chB (Fig. 5A and C) and a very well defined electron density. The RMSD of the loop is lower even than the divergence of any of the helices (compare lo and hI, hII, hIII in Table 6). Similarly to the N-terminal, arm the flexibility of the loop of WUS HD is restricted by two of its amino acids, which connect it firmly with the helix bundle. Since the hydrophobic core plays a central role, docking of the N-terminal arm and the loop will be described in more detail in section 3.1.5 after the amino acid residues, which compose the hydrophobic core, are identified.

Table 6: Mean distance of corresponding structural elements of chA and chB - helix I, II and III (h I, h II, h III), N-terminal arm (Na), loop (lo) and turn (tu) - determined for C α -atoms (C α) and all atoms (all).

		Na	h I	lo	h II	tu	h III
C α	no. of atoms	5	12	8	14	4	13
	RMSD [Å]	1.55	0.21	0.07	0.10	0.24	1.36
all	no. of atoms	46	108	60	118	30	117
	RMSD [Å]	3.67	0.82	0.19	0.33	0.59	1.53

3.1.3 Hydrogen bonds and electrostatic interactions

Homeodomains often display several salt bridges and hydrogen bonds stabilizing the tertiary structure. Comparably frequent are interactions between charged side chains of the residues

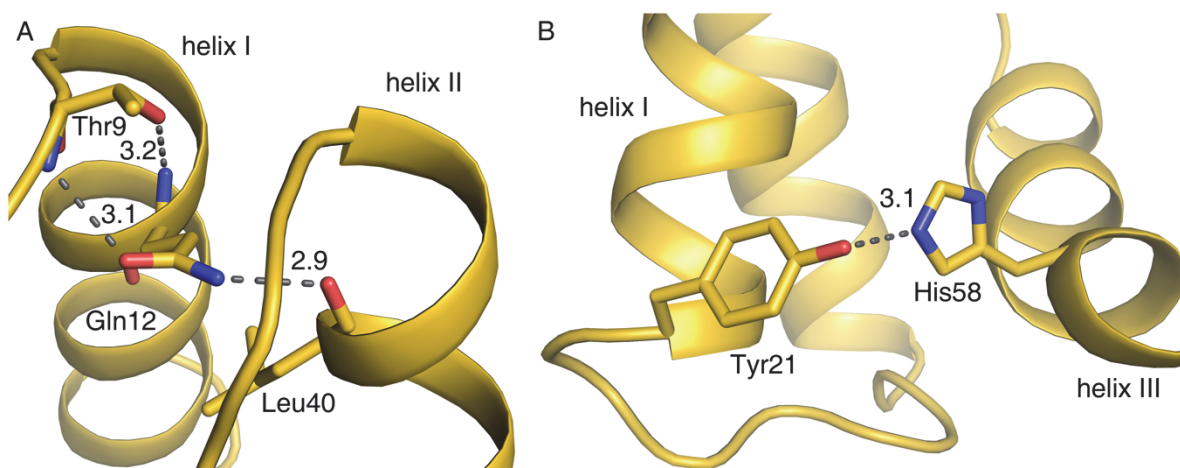


Fig. 6: Intramolecular hydrogen bonds in WUS HD. A) Network of hydrogen bonds centred on Gln12. B) Hydrogen bond between side chains of Tyr21 and His58.

Glu17/Arg52 and Glu19/Arg30 (Clarke 1995) and hydrogen bonds between the pairs Gln12/residue 38 and Arg53/residue 24, which connect the side chains of Gln12 and Arg53 with backbone carbonyl groups of the paired up residue (Clarke, Kissinger et al. 1994; Baird-Titus, Clark-Baldwin et al. 2006) (residue numbering according to typical homeodomains). Of those

four common intramolecular contacts Gln12 is the only one present in WUS HD. The side chain carbonyl oxygens of Thr9 and Gln12 mutually interact with each others' backbone amides (Fig. 6A). The side chain amide of Gln12 donates its hydrogen to the carbonyl oxygen of the Leu40 main chain and thus links the N-terminus of helix I with the C-terminus of helix II. An additional but less frequently detected hydrogen bond is established between the terminal functional groups of Tyr21(OH) and His58(Ne2) side chains and connects the C-termini of helices I and III in WUS HD (Fig. 6B).

3.1.4 The hydrophobic core

The accessible surface area (ASA) of each residue and atom was determined in ccp4mg (McNicholas, Potterton et al. 2011) with a water radius of 1.4 Å and the point density set to 1 Å. The ASA of atoms belonging to the side chain were summarised (scASA) and normalized to their respective maximum surface area (Miller, Janin et al. 1987) to obtain relative scASA (% scASA). Residues with values below 16 % in at least one of the two WUS HD molecules are listed in Table 7 below.

Table 7: Amino acid residues of WUS HD, whose side chains are accessible to solvent for less than 16 % of the total surface area (% scASA); % scASA values lower than 10 % (yellow) or between 10 % and 16 % (orange) are highlighted; calculations were performed on chA, chB and chA*, a truncated version of chA comprising residues Arg5-Glu62; “x” indicates the absence of a corresponding residue

% scASA groups	residue		% scASA		
			chA	chA*	chB
<0.6%	12	Gln	0.10	0.10	0.47
	16	Leu	0.57	0.57	0.26
	40	Leu	0.04	0.04	0.04
	46	Ile	0.19	0.19	0.14
	51	Val	0.17	0.26	0.03
	55	Phe	0.41	0.41	0.40
3%	19	Leu	6.66	6.66	7.45
-	28	Pro	2.86	2.86	2.82
10%	36	Ile	7.57	7.57	7.93
	37	Thr	6.49	6.49	6.97

% scASA groups	residue		% scASA		
			chA	chA*	chB
10%	6	Trp	2.30	13.37	15.80
	15	Ile	15.34	15.34	17.84
-	20	Tyr	12.59	12.59	11.04
16%	25	Ile	12.95	12.95	13.25
	54	Trp	6.80	9.76	13.71
>16%	58	His	16.45	16.45	24.52
(first 2 res.)	33	Ile	23.17	23.17	25.09
artefacts	2	Thr	9.69	x	x
	8	Pro	6.28	38.05	32.87
	50	Asn	7.44	21.93	20.59
	61	Arg	12.28	40.59	x

Most of the discrepancy between the two WUS HD copies, chA and chB in Table 7, can be explained by differences in the N-terminal arm of which fewer residues are resolved in chB. In chA it is placed over the turn and runs anti-parallel to helix III. On its way several amino acid residues are partially buried beneath, whose relative scASA values increase when the calculation is performed on a truncated version of chA missing residues 1-4 (Table 7, compare chA

and chA*). As is described in section 3.1.2, the N-terminal arm of chA is stabilized by intermolecular contacts in the crystal and does not necessarily represent its native location. Therefore subsequent assessment of core amino acid residues will only take calculations for chB and the truncated version of chA (chA*) into account. Amino acid residues, which are mostly buried by residues 1 to 4 are considered artefacts.

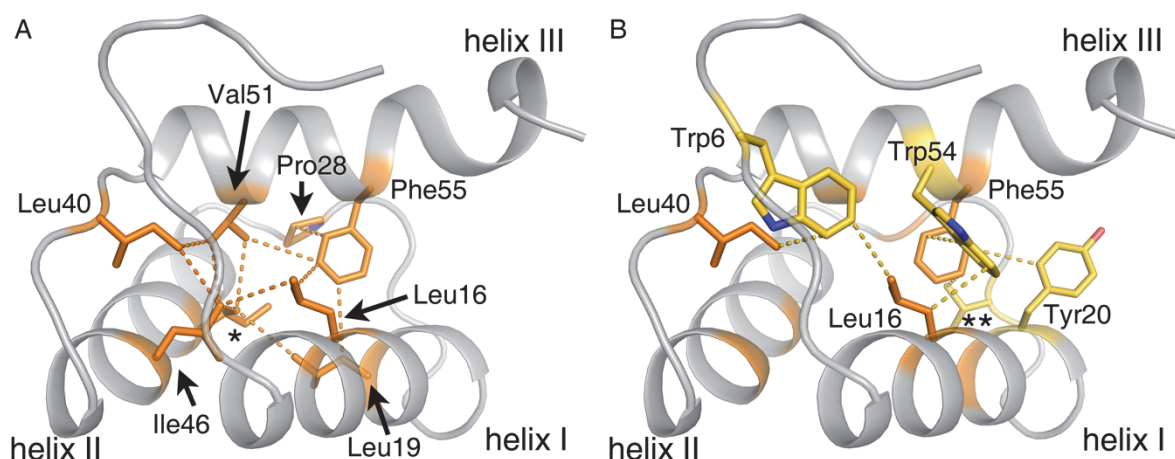


Fig. 7: Amino acid residues forming the hydrophobic inner core and outer core of WUS HD. A) and B) chA of WUS HD shown as ribbon; amino acid side chains of hydrophobic inner core (orange) and outer core (yellow) are displayed as sticks; hydrophobic interactions are depicted as dashed lines, distances range in between 3.6 Å and 4.1 Å. A) The hydrophobic inner core (orange) in WUS HD comprises the residues Leu16, Leu19, Pro28, Ile36 (*), Leu40, Ile46, Val51 and Phe55, which form a tight network. B) Residues Trp6, Tyr20, Ile25 (**) and Trp54 of the outer core (yellow) interact with amino acids of the inner core (orange).

The two residues Gln12 and Thr37 display polar amino acids and therefore are not regarded as part of the hydrophobic core despite being deeply buried within the protein (Table 7). Hydrophobic amino acid residues of WUS HD with relative scASA values below 10 % establish a tight network (Fig. 7A). In the list of relative scASA they are followed by a cluster of residues with mostly amphiphilic side chains, which includes Tyr20 and Trp54 (Table 7). More detailed analysis of their location and interacting partners in the WUS HD structure revealed that all amino acid residues with relative scASA values between 10 % and 16 % contact at least one of the deeply buried amino acids with values below 0.6 % (Table 7 and Fig. 7B), whereas amino acid residues above the threshold of 16 % scASA - e.g. His58 and Ile33 - do not (data not shown). However, some facts indicated the need of an additional subdivision. First the presence of hydrophobic amino acids (amino acids with positive hydrophathy values according to Kyte and Doolittle (Kyte and Doolittle 1982)) diminishes strongly above a value of 10 % relative scASA and amino acids exhibiting amphiphilic side chains, e.g. tryptophan or tyrosine, are more frequent (Table 7). Additionally the relative accessibility to solvent for equivalent residues in

the two WUS HD copies start to diverge extensively above the threshold of 10 %, whereas below this threshold the relative scASA values are nearly identical for the two chains (Table 7). This suggests an increase in flexibility and higher degree of freedom for the according side chain, which is unexpected of the usually tightly packed protein core. For these reasons the threshold of 10 % relative scASA was selected to differentiate the hydrophobic core between amino acid residues of the inner and outer core.

Summarizing the section above, the hydrophobic core of WUS HD is formed by residues Trp6, Leu16, Leu19, Tyr20, Ile25, Pro28, Ile36, Leu40, Ile46, Val51, Trp54 and Phe55 (Table 7, Fig. 7A and B), which may be grouped into amino acid residues of the inner and of the outer core. The inner core comprises residues Leu16, Leu19, Pro28, Ile36, Leu40, Ile46, Val51, Phe55, which form a tight hydrophobic network (Fig. 7A). The inner core is surrounded by residues of the outer core Trp6, Tyr20, Ile25 and Trp54, which do not interact with each other but contact amino acids of the inner core (Fig. 7B), are more accessible to solvent and display a higher degree of freedom than amino acid residues of the inner core (Table 7)

3.1.5 Stabilization of the N-terminal arm and the loop in WUS HD

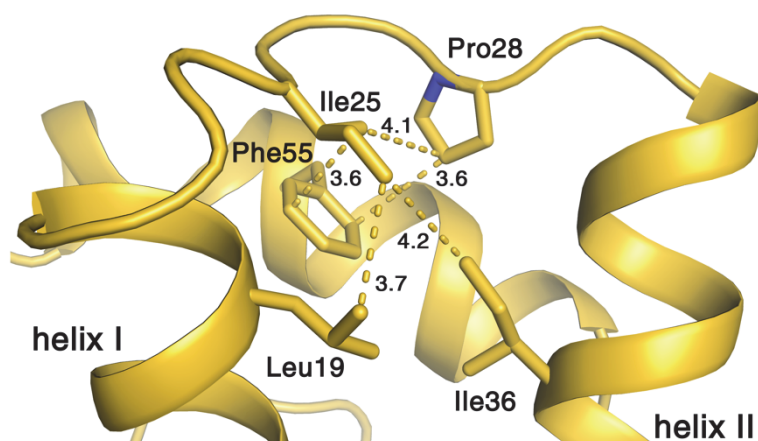


Fig. 8: Hydrophobic network established by Ile25 and Pro28. Interactions are indicated by dashed lines and distances are given in [Å].

As mentioned above, the loop of WUS HD (residues 22 to 29; Fig. 4A) seems to be highly rigid indicated by the extremely low divergence of the the two copies of WUS HD in this region (section 3.1.2; Fig. 5A and C; Table 6) and a very well-defined electron density. Two amino acid residues of the loop were identified to be part of the hydrophobic core, Ile25 and Pro28 (section 3.1.4). The proline residue and an extensive hydrophobic network established by Ile25 appear to reduce the flexibility of the loop to a minimum (Fig. 8).

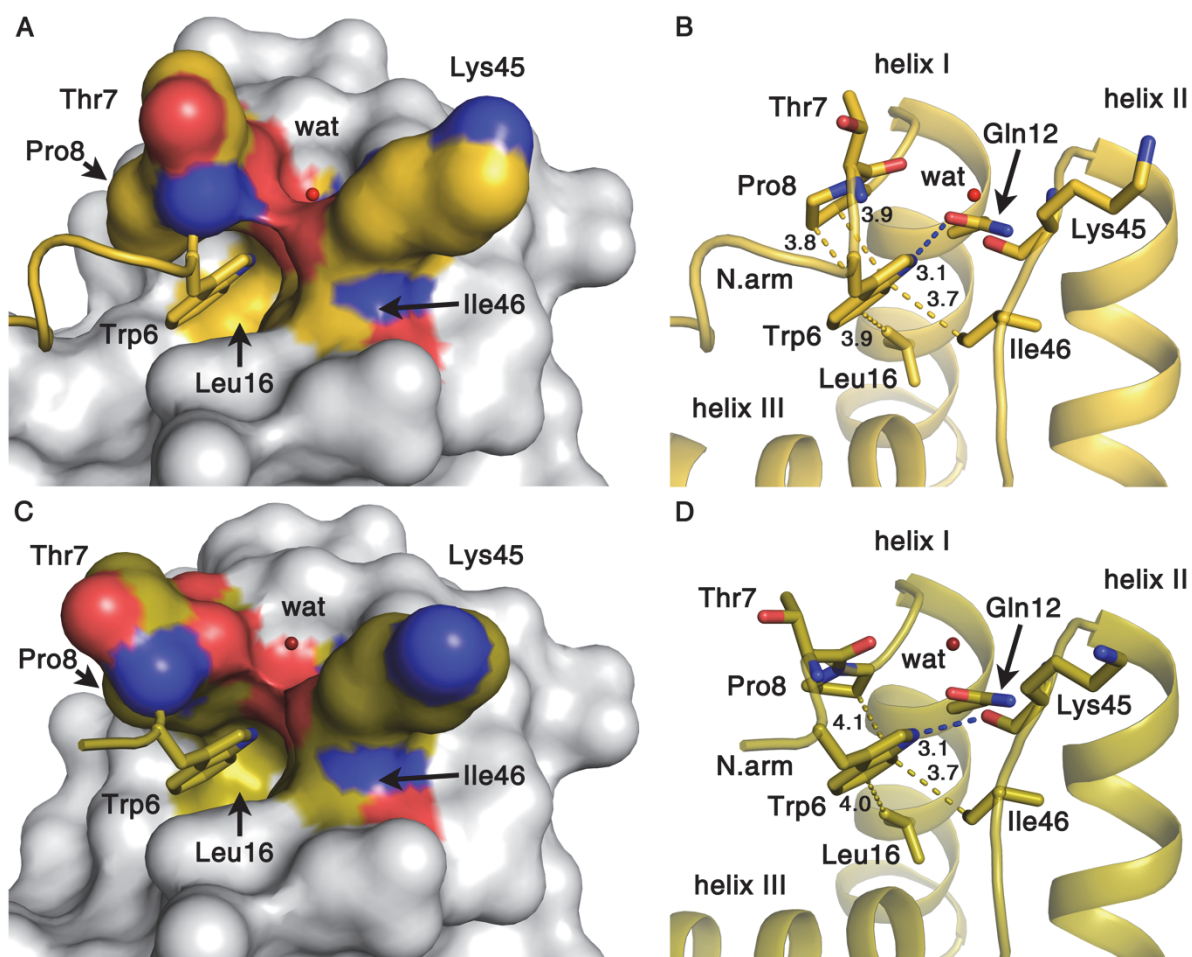


Fig. 9: Docking of the N-terminal arm to the helix bundle. The Trp6-pocket of WUS HD in chA (A and B) and chB (C and D). A) and C) surface is displayed for all residues C-terminal of Trp6; side chain of Trp6 is depicted as stick. B) and D) WUS HD displayed as ribbon; Trp6 and residues forming the pocket to fit Trp6 are shown as sticks; molecular interactions are indicated by dashed lines and distances are given in [Å]. WAT: Coordinated water molecule depicted as red ball. N.arm: N-terminal arm

Interest in the amino acid residue Trp6 was already awakened by the NCS of the two WUS HD copies in the crystal (section 3.1.2). Though the conformation of the N-terminal arm of chA and chB differ strongly, a sudden incision is apparent at residue Trp6 (Fig. 5C). In particular the positions of the Trp6 side chains of the two copies are nearly identical (Fig. 5A). In section 3.1.4 Trp6 was detected as part of the hydrophobic core and thus links the N-terminal arm with the inner center of the helix bundle. The Trp6 side chain is inserted into a pocket, which appears as the perfect fit for the amino acid tryptophan. The side chains of the residues Pro8, Leu16 and Ile46 form the hydrophobic portion of the pocket, which is spanned by a group of carbonyl oxygens of the main chains of Thr7 and Lys45 and of the side chain of Gln12 and a coordinated water molecule (Fig. 9A, C). Hydrophobic fractions of the inserted Trp6 side chain are enclosed by Pro8, Leu16 and Ile46, whereas the NH-group of the Trp6 indole ring establishes a polar

contact with the carbonyl group of the Gln12 side chain in chA and that of the Lys45 backbone in chB, respectively (Fig. 9B, D). Thus the side chain of Trp6 is nearly entirely encaged with minimal degree of freedom resulting in the curious effect that the main chain of an amino acid residue is more flexible than its side chain.

3.2 Comparative analysis of WUS HD and other homeodomain structures

Five homeodomains representing the three basic structural types of homeodomains, (typical, TALE and other atypical) from basal and derived species were selected for comparative analysis with the WUSCHEL homeodomain: the yeast homeodomains MATING TYPE a1 (MATa1 HD, typical) and α 2 (MAT α 2 HD, TALE), the homeodomains ENGRAILED (EN HD, typical) and EXTRADENTICLE (EXD HD, TALE), both originating from *Drosophila melanogaster*, and the atypical homeodomain of the human HEPATOCYTE NUCLEAR FACTOR 1 β (HNF1 β HD, atypical). Except for one of the two EN HD structures all homeodomains were co-crystallized with DNA. For more details refer to Table 5 in section “Materials and Methods”. Individual residues are numbered according to the respective PDB file (file format of the structures deposited in the Protein Data Bank).

3.2.1 Identification of amino acid residues of WUS HD without structural equivalent in typical and atypical homeodomains

Pairwise structural alignment to WUS HD utilizing the webtool RAPIDO (Mosca, Brannetti et al. 2008; Mosca and Schneider 2008) revealed a highly conserved backbone structure of corresponding residues (Fig. 10 A-C). The mean distance (root mean square deviation, RMSD) of C α -atoms ranges in between 1.27 Å and 1.74 Å with MATa1 HD being the most similar and HNF1 β HD the most diverging structure (RMSD rig in Table 8).

When the threshold for allowed divergence (Low Limit or LoLim) was lowered to 1.5 Å for typical and TALE homeodomains and to 1.0 Å in case of HNF1 β HD, respectively, RAPIDO inserted at least one break into the structure splitting it into so called rigid bodies. The rigid bodies are then aligned separately to WUS HD and the RMSD is calculated as the difference score of corresponding C α -atoms summarized over all individual rigid bodies (RMSD flex). The flexible superimposition facilitates the alignment of a higher number of amino acid residues and led to a decrease in structural divergence (Table 8), which is also apparent in the neat superimposition of the homeodomain structures (Fig. 10 D-F).

Table 8: Statistical analysis of the pairwise structural alignments of selected homeodomains to WUS HD; PDB ID: accession number of the structure deposited in the Protein Data Bank.

	HD of (PDB ID)	MATa1 (1YRN)	EN (1ENH)	MAT α 2 (1APL)	EXD (1B8I)	HNF1 β (2H8R)
	no. of residues in total	49	54	59	58	78
rigid superimposition	no. of aligned residues	46	40	41	41	37
	RMSD rig [Å]	1.27	1.56	1.7	1.53	1.74
flexible superimposition	no. of aligned residues	49	53	50	50	57
	RMSD flex [Å]	0.64	0.6	0.56	0.55	0.53
	no. of rigid bodies	3	2	2	2	3

Despite the high structural similarities between WUS HD and the other examined homeodomains, direct comparison and calculation of RMSD is feasible for corresponding residues only. However, in contrast to the homeodomain of WUSCHEL, which comprises 66 amino acids, typical and TALE homeodomains consist of only 60 and 63, respectively. Thus at least in case of six or three amino acid residues of WUS HD it is technically impossible to assign a corresponding residue, which are referred to as gapped or unaligned amino acid residues (magenta, Fig. 10). Please note the difference between gapped (magenta and red in Fig. 10) and flexibly aligned amino acid residues (dark blue in Fig. 10). In contrast to gapped amino acid residues, flexibly aligned amino acid residues are matched to a equivalent in the other homeodomain, even though only rather roughly.

As consequence of the technical impossibility to assign a corresponding amino acid residue to all homeodomain positions of WUS HD, some regions of its structure have to diverge in conformation. In general differences in homeodomain structure are referred to as in comparison to typical homeodomains only. However, it has been suggested that the additional amino acids of atypical homeodomains may always be inserted in similar regions between the helices (reviewed in Bürglin 1994). So in order to identify structural characteristics unique to WUS HD, it seemed reasonable to not only compare it to typical homeodomains but also to other atypical homeodomains. Apart from the termini of WUS HD two regions are detectable, which are composed of amino acids without structural equivalent in the other examined homeodomains (magenta, Fig. 10). In the following these regions will be referred to as Divergent Region I and II or DR I and II. One of them is located in the loop, which connects helix I and II (DR I), the other at the C-terminus of helix II and adjacent residues of the turn (DR II). Though the conformational changes seem to be the consequence of the integration

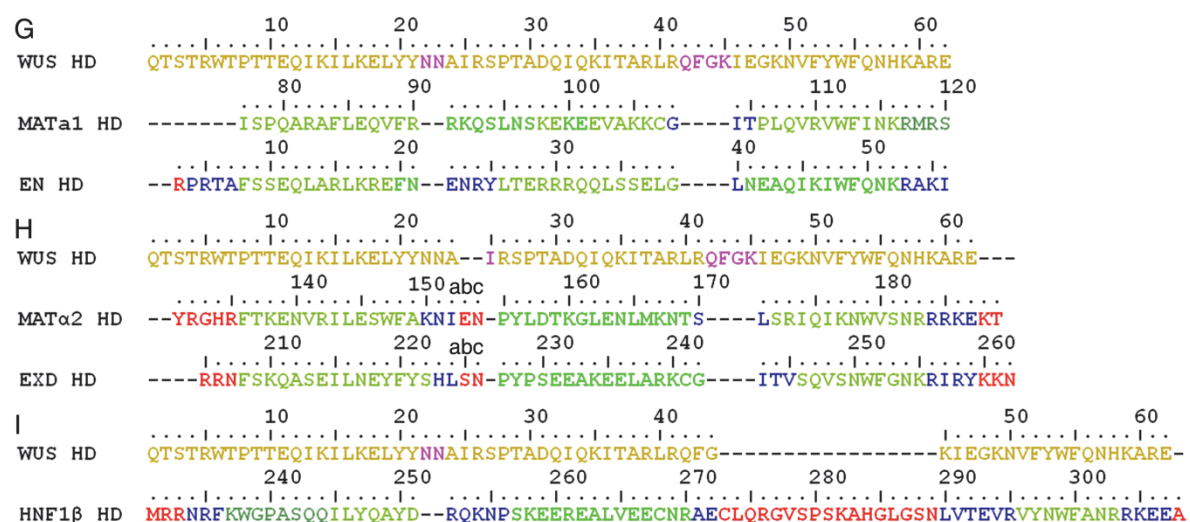


Fig. 10 (previous page): Comparison of WUS HD to three different types of homeodomain structures. A-F) Superimposed structures of WUS HD (yellow) with the typical homeodomains of MATa1 and EN (A and D), the TALE homeodomains of MATa2 and EXD (B and E) and the atypical homeodomain of HNF1 β (C and F). A-C) Rigid superimposition. D-F) Flexible superimposition of WUS HD and the individual rigid bodies (RB). G-I) Amino acid sequence alignments in accordance with the results obtained by the structural alignments depicted in A-F. RB: different shades of green; flexibly aligned amino acid residues: dark blue; gapped, unaligned amino acid residues: magenta in WUS HD, red in the other homeodomain structure; hI, hII, hIII: helices I, II and III.

of additional amino acids, this study discriminates between homeodomain positions displaying a divergent backbone structure (DR I and II) and the actual extra amino acid residues.

The results given for the gapped amino acid residues of DR I by the webtool RAPIDO differ depending on which structural type of homeodomain WUS HD is compared to. In case of typical homeodomains, represented by MATa1 HD and EN HD (Fig. 10 A, D and G) in this study, but also in case of HNF1 β HD (Fig. 10 C, F and I) the program can not assign a structural equivalent to residues Asn22 and Asn23 of WUS HD. However, when superposed with TALE homeodomains, the gap is shifted slightly towards the C-terminus. In these cases RAPIDO flexibly aligns Ala24 to “a”, one of the additional amino acids of TALE homeodomains, and inserts a gap at the position corresponding to Ile25 (Fig. 10 B, E and H). Thus the comparison of WUS HD to typical or TALE homeodomains identifies Asn22 and Asn23 or Ile25 as DR I.

When superimposed with typical and TALE homeodomains a second divergent region (DR II) is apparent in WUS HD, which is caused by the integration of the remaining four of the six extra residues of WUS HD. In these cases amino acids of WUS HD without structural equivalents include the successive residues Gln42 to Lys45, which thus comprise DR II (Fig. 10 A, B, D, E, F and G). In the structural alignment of the atypical homeodomain of HNF1 β and WUS HD, however, all residues of DR II can be assigned an equivalent in the other homeodomain. Arg270 of HNF1 β HD corresponds to Gln42 of WUS HD (Fig. 10 C, F and I) and the remaining amino acids of DR II, Phe43 to Lys45, are at least flexibly aligned to residues 271, 272 and 290 of HNF1 β HD (Fig. 10 I).

Thus the regions subjected to conformational changes, meaning residues without structural equivalent in typical homeodomains, seem roughly similar between WUS HD and other atypical homeodomains. Amino acids of DR I identified by comparison with typical homeodomains overlap partially with the additional amino acids of TALE homeodomains and amino acids of DR II can be aligned to fractions of the extended regions of HNF1 β HD. However, in contrast to the three additional amino acids of TALE HD and the twentyone of HNF1 β HD, DR I and II are composed of only two and four amino acids, respectively. Also the changes in conformation

display obvious differences between the situation of WUS HD and that of the other examined atypical homeodomains. In case of DR I a hydrophobic core amino acid residue, which is unique to WUS HD, is essential for the conformation of DR I. Hence a more detailed analysis of DR I and also of DR II will be given in section 3.2.3.2 and section 3.2.3.3 after the comparison of core amino acid residues.

In the following a multiple sequence alignment will be used, which is based on the pairwise structural alignment shown in Fig. 10. But slight modifications of the original sequence alignment shown in Fig. 10 G-I were necessary. TALE homeodomains contain three additional amino acids labelled “a”, “b” and “c” inserted in between the residues 23 and 24 of typical homeodomains (Wolberger, Vershon et al. 1991). The first one, “a”, of both analysed TALE homeodomains was designated to be the corresponding residue of Ala24 of WUS HD (Fig. 10 H). When compared to the two typical homeodomain structures Ala24 of WUS HD corresponds to Arg91 in MATa1 HD and Glu22 in EN HD (Fig. 10 G). In the multiple sequence alignment of all analysed homeodomain this would lead to the impression that “a” has a structural equivalent in typical homeodomains, which is not the case by definition. So in contrast to the original results shown in Fig. 10, alignment of Ala24 of WUS HD to “a” of TALE homeodomains will be neglected in case of the multiple amino acid sequence alignments shown in subsequent sections.

3.2.2 Comparison of core amino acid residues

The same method used to determine buried amino acid residues of WUS HD was adopted in the case of the other five homeodomains examined in this study. Except for EXD HD, for which no second complete structure was available, two structures of each homeodomain were analyzed to assess their intrinsic variability. Homeodomains exhibit a highly conserved hydrophobic core formed by amino acid residues 16, 20, 26, 34, 35, 38, 40, 45, 48 and 49 (numbering according to typical homeodomains, Fig. 1, Fig. 2), all of which are readily identified as core amino acid residues in the homeodomain structures examined in this study (Table 9, Fig. 11) supporting the reliability of the method applied. Given the high structural similarities of homeodomains in general it is not very surprising that the hydrophobic core of is also conserved in WUS HD. Nonetheless divergencies are noticeable at positions 6, 12, 19 and 25 of WUS HD (Fig. 11) and require a more detailed examination. In the following labelling of amino acid residues is according to the labels of the respective PDB file. If present, the number of the corresponding

Table 9: Accessible surface area (ASA) in [%] of the side chains (sc) of amino acid residues in selected homeodomain structures normalized to the respective maximum ASA determined by Miller and colleagues (Miller, Janin et al. 1987). Values are shown for amino acid residues, of which at least one is lower than 16 % as well as the values of the corresponding positions in the other homeodomains. Values below 10 % (orange) and between 10 % and 16 % (yellow) are highlighted. x: either no corresponding amino acid residue is present or it is a glycine residue; chA*: amino acid residues N-terminal of Arg5 were cut for this analysis; labelling is according to the respective pdb file; for the position number of corresponding amino acid residues in typical homeodomains, please be referred to the structures of EN HD.

WUS HD			MATa1 HD			EN HD			MATa2 HD			EXD HD		HNF1β HD		
	chA*	chB		chA 1YRN	chA 1AKH		chA* 1ENH	chA 3HDD		chC 1APL	chB 1YRN		chA 1B8I		chA 2H8R	chB 2H8R
6 TRP	13.4	15.8	x	x	x	6 THR	47.9	62.2	134 HIS	25	27.3	206 ARG	72.1	236 PHE	9.7	10.7
8 PRO	38	32.9	77 ILE	24.8	24	8 PHE	4	17.8	136 PHE	23.9	25.1	208 PHE	23.8	238 TRP	16.3	15.6
11 GLU	54.5	50.2	80 GLN	77.5	77	11 GLU	101	94.9	139 GLU	68.6	81.6	211 GLN	67.2	241 ALA	13.5	24.5
12 GLN	0.1	0.5	81 ALA	0.9	0.5	12 GLN	21.9	19.2	140 ASN	7.6	20.4	212 ALA	1.4	242 SER	3.2	3
15 ILE	15.3	17.8	84 PHE	35.3	35	15 ARG	32.8	24.7	143 ILE	23.8	21.7	215 ILE	26.8	245 ILE	18.9	13.9
16 LEU	0.6	0.3	85 LEU	0	0	16 LEU	0.1	0.8	144 LEU	0.9	0.2	216 LEU	0	246 LEU	0.9	0.2
19 LEU	6.7	7.5	88 VAL	18.5	17.5	19 GLU	23.3	14.8	147 TRP	16.2	13.5	219 TYR	18.3	249 ALA	8.6	24.3
20 TYR	12.6	11	89 PHE	14	12.2	20 PHE	6.6	10.6	148 PHE	4.4	2.8	220 PHE	1.9	250 TYR	24.2	17.3
25 ILE	13	13.3	92 LYS	53	48.5	23 ASN	57.2	69.5	x x	x	x	x x	x	253 GLN	53.4	40.8
26 ARG	54.9	47.2	93 GLN	83.4	64.1	24 ARG	48.9	67.9	155 PRO	12.4	18.4	227 PRO	16.9	254 LYS	95.5	66
28 PRO	2.9	2.8	95 LEU	5.5	4.9	26 LEU	6.5	11.6	157 LEU	8.8	6.9	229 PRO	0.6	256 PRO	5.2	10.4
33 ILE	23.2	25.1	100 LYS	34	33	31 ARG	22.3	29.8	162 LEU	15.1	10.1	234 LYS	17	261 ARG	21.1	17
36 ILE	7.6	7.9	103 VAL	10.1	8.6	34 LEU	3.2	0.2	165 LEU	5	3.7	237 LEU	9.2	264 LEU	11.2	14.5
37 THR	6.5	7	104 ALA	4.1	5.1	35 SER	20.7	23.6	166 MET	33.3	25.7	238 ALA	8.8	265 VAL	5.1	16.7
39 ARG	52.9	57.4	106 LYS	69.2	71.9	37 GLU	53.4	50.3	168 ASN	54.1	59.7	240 LYS	15.1	267 GLU	55.1	60.7
40 LEU	0	0	107 CYS	0.1	0.4	38 LEU	8.8	9.9	169 THR	0	0	241 CYS	0	268 CYS	0.5	0.3
44 GLY	x	x	x	x	x	x	x	x	x	x	x	x	x	272 GLU	6.2	4
x	x	x	x	x	x	x	x	x	x	x	x	x	x	273 CYS	1.9	1.8
x	x	x	x	x	x	x	x	x	x	x	x	x	x	283 ALA	13.5	20
x	x	x	x	x	x	x	x	x	x	x	x	x	x	286 LEU	3.1	5.1
46 ILE	0.2	0.1	109 ILE	6.6	7.9	40 LEU	6	8.2	171 LEU	6.5	14.2	243 ILE	5.5	291 VAL	0.6	1.3
50 ASN	21.9	20.6	113 GLN	46.6	47.1	44 GLN	28.5	24.9	175 GLN	39.8	30.4	247 GLN	51	295 ARG	39.8	37.3
51 VAL	0.3	0	114 VAL	0.8	0.3	45 ILE	0	0.1	176 ILE	0.1	0	248 VAL	0	296 VAL	0.9	1.4
54 TRP	9.8	13.7	117 TRP	17.2	15.3	48 TRP	2.4	13.4	179 TRP	16.9	16	251 TRP	15.3	299 TRP	3.9	5.6
55 PHE	0.4	0.4	118 PHE	3.9	3.6	49 PHE	1.2	2.1	180 VAL	0.3	0.5	252 PHE	1.1	300 PHE	2.2	4.6
58 HIS	16.5	24.5	121 LYS	26.6	23.1	52 LYS	20.1	24.1	183 ARG	15.5	17.3	255 LYS	15.4	303 ARG	23.3	22.9

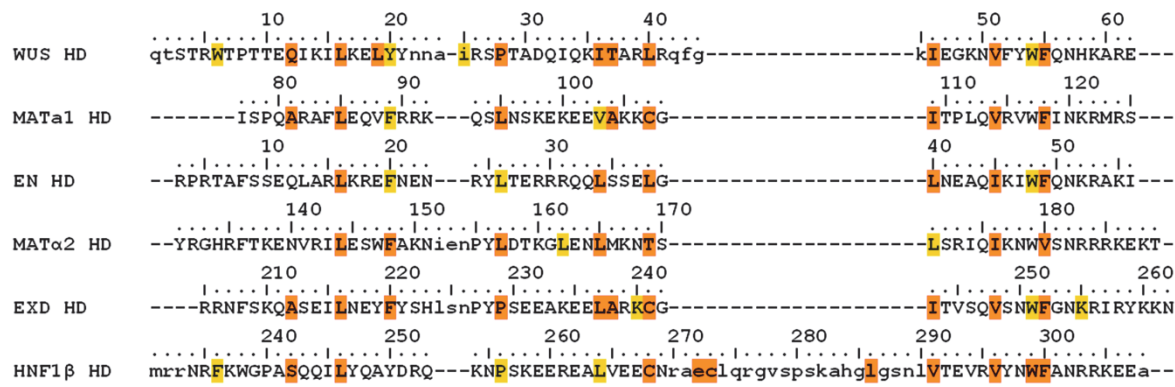


Fig. 11: Comparison of core amino acid residues in selected homeodomains. Amino acid residues with relative scASA values below 10 % (orange) and between 10% and 16 % (yellow) in both examined structures of the respective homeodomain (Table 9) are highlighted.

position in typical homeodomains is given in parenthesis after the amino acid residue, e.g. Leu157 (typ26).

First subject of analysis is Leu19 (typ19) of WUS HD, a position of the homeodomain structure better known as being part of a salt bridge (Glu19/Arg30) especially in HOX homeodomains (Clarke 1995). In MATA1 HD Val88 (typ19), which corresponds to Leu19 of WUS HD, serves as part of the hydrophobic contact surface for dimerization with MATA2 (Li, Stark et al. 1995). Consequential the artificial removal of MATA2 from the heterodimeric structure renders much of the Val88 (typ19) side chain accessible to solvent (Table 9), but does not necessarily display the situation of the native MATA1 monomer. In general relative scASA values for amino acids corresponding to Leu19 (typ19) of WUS HD seem to vary greatly even between two structures of the same homeodomain (Table 9). A big portion of the surface of Leu19 (typ19) in WUS HD is hidden by the side chains of Asn23 and Arg39 (typ37) (Fig. 12 A). Asn23 is part of DR I and the reach of its side chain is extended by the unusual conformation of DR I in WUS HD. In this way the terminal end of the comparably short side chain of Asn23 is placed in a position similar to Arg252 (typ22), which shelters Ala249 (typ19) in HNF1β HD (chA) from solvent (Fig. 12 A). Also the difference in accessibility of the corresponding amino acid residue in the two EN HD structures, Gln19, is largely due to diverging conformations of two side chains, which cover Glu19 in one structure but do not in the other (Fig. 12 B). Thus, the amino acid at position 19 seems highly variable in general and Leu19 (typ19) may not be considered a unique core amino acid residue of WUS HD. Since Gln12 (typ12) is covered by DR II (Fig. 13; for more details on DR II see section 3.2.1), it is sort of a unique core amino acid residue of WUS HD. However, given the nearly indistinguishable conformations and interactions of Gln12 in WUS HD and in EN HD (Fig. 13), it does seem

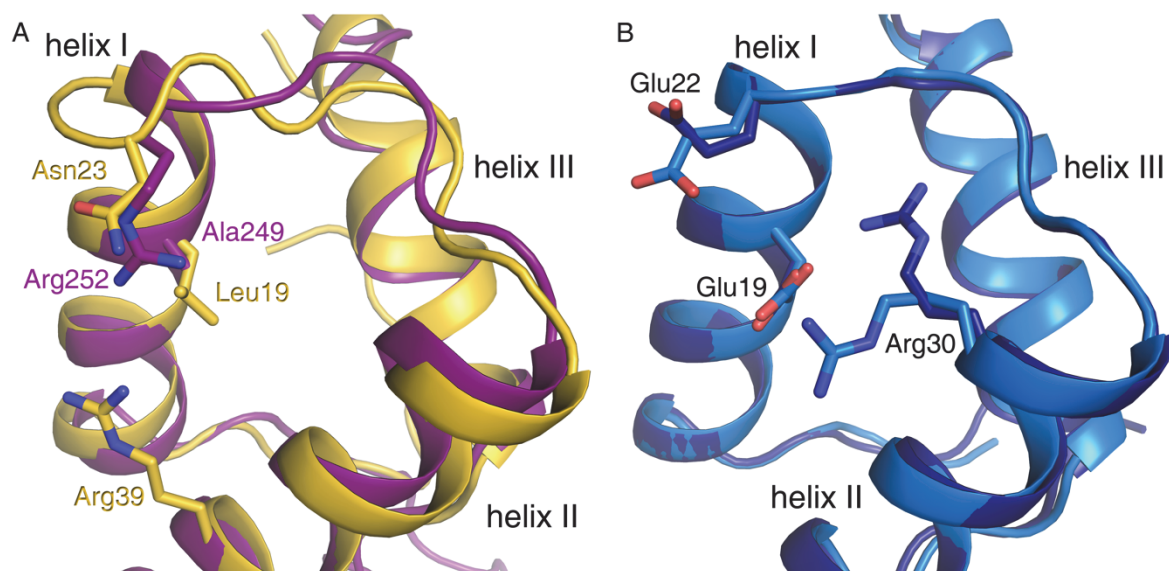


Fig. 12: Solvent accessibility of Leu19 depends on conformation of adjacent side chains. A) Superimposed structures of WUS HD (yellow) and HNF1 β HD (purple); B) Superimposed structures of two EN homeodomains with the PDB IDs 1ENH (dark blue) and 3HDD (marine). Homeodomains are displayed as ribbons; Residues Leu19 (typ19), Ala249 (typ19), Glu19 and side chains covering them are depicted as sticks.

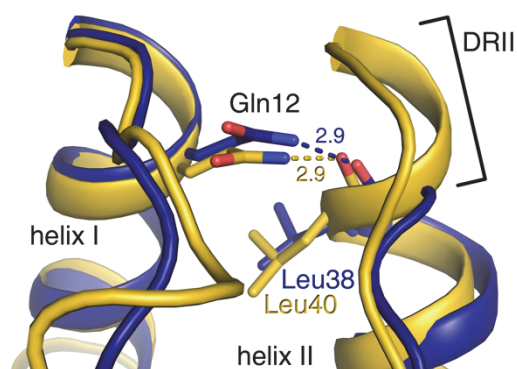


Fig. 13: Coverage of Gln12 in WUS HD and EN HD. Superposed structures of EN HD (blue) and WUS HD (yellow) displayed as ribbons; Gln12 in both structures as well as the corresponding residues Leu38 in EN HD and Leu40 in WUS HD are depicted as sticks. Residues ranging from Gln42 to Lys46 (bracket) of DR II in WUS HD (see section 3.2.1) shelter Gln12 from the solvent.

unlikely that DR II and the unusual high coverage of Gln12 (typ12) apparent in the WUS HD structure are evolutionary linked, but rather appear to be a coincidence.

Two amino acid residues with unusual low accessibility to solvent in WUS HD remain to be investigated, Trp6 (typ6) and Ile25 (Fig. 11). Both function in stabilizing flexible regions of the WUS HD structure (see section 3.1.5). Moreover in case of Ile25 assigning a distinct structural equivalent amino acid residue in the other examined homeodomains proved to be difficult (see section 3.2.1), which already indicated that Ile25 is a potential structural characteristic of WUS HD. Due the need of integrating results of other sections in order to properly describe and

compare the function of Trp6 (typ6) and Ile25 a more detailed analysis is given in sections 3.2.3.1 and 3.2.3.2.

3.2.3 Novel structural characteristics of WUS HD

3.2.3.1 Stabilization of the N-terminal arm distinguishes WUS HD from other published homeodomain structures

It is striking that some of the most conserved amino acids in homeodomains are altered in WUS HD. One of them is the amino acid residue 8, most frequently a phenylalanine or tyrosine (Fig. 1), which docks the N-terminal arm to the helix bundle (Noyes, Christensen et al. 2008). Homeodomains lacking a large hydrophobic amino acid at residue 8 appear unable to stably interact with the minor groove of the DNA, which is usually facilitated by the N-terminal arm (Li, Stark et al. 1995; Noyes, Christensen et al. 2008) (see also section 1.1.3.2 and Fig. 3). In WUS HD, however, Phe/Tyr8 is exchanged for a proline residue, Pro8 (typ8) (Fig. 4 A), and the N-terminal arm is stabilized by Trp6 (typ6) instead (section 3.1.5). Interestingly, the side chain of Trp6 (typ6) extends into a similar space as the side chain of Phe/Tyr8 in other homeodomains exemplified by Phe208 (typ8) and Phe8 (typ8) in Fig. 14 A. Also the interactions between Trp6 (typ6) and the conserved hydrophobic core amino acid residues Leu16 (typ16) (distance of 3.9 Å) and Ile46 (typ40) (distance of 3.7 Å) are essentially the same as those formed by Phe/Tyr8 in the other examined homeodomains (Fig. 14 A), e.g. in EXD HD the distance of Phe208 (typ8) to Leu216 (typ16) is 4.0 Å and to Ile243 (typ40) 3.7 Å. However, in contrast to Phe/Tyr8 apparent in most other homeodomains, Trp6 (typ6) interacts with two additional amino acid residues in the structure of WUS HD. One is an extra hydrophobic interaction with Pro8 (typ8) (3.8 Å) and the other one is established between the NH-group of the Trp6 (typ6) indole ring and the carbonyl oxygen of the Gln12 (typ12) side chain in chA or Lys45 (DR II) main chain in chB, respectively (Fig. 14 B; for more details see section 3.1.5). As a result, the side chain of Trp6 (typ6) is enclosed almost entirely (Fig. 15 B) with only very little of its surface accessible to solvent (Table 9), which is in contrast to Phe/Tyr8 (typ8), exemplified by Phe208 (typ8) of EXD in Fig. 15 C. Of the amino acids in the other examined homeodomains at positions corresponding to Trp6 (typ6) only Phe236 (typ6) in HNF1β HD exhibited properties partially resembling Trp6 (typ6) in WUS HD and thus with the potential of taking on a similar function. Both, Trp6 (typ6) and Phe236 (typ6) exhibit large hydrophobic

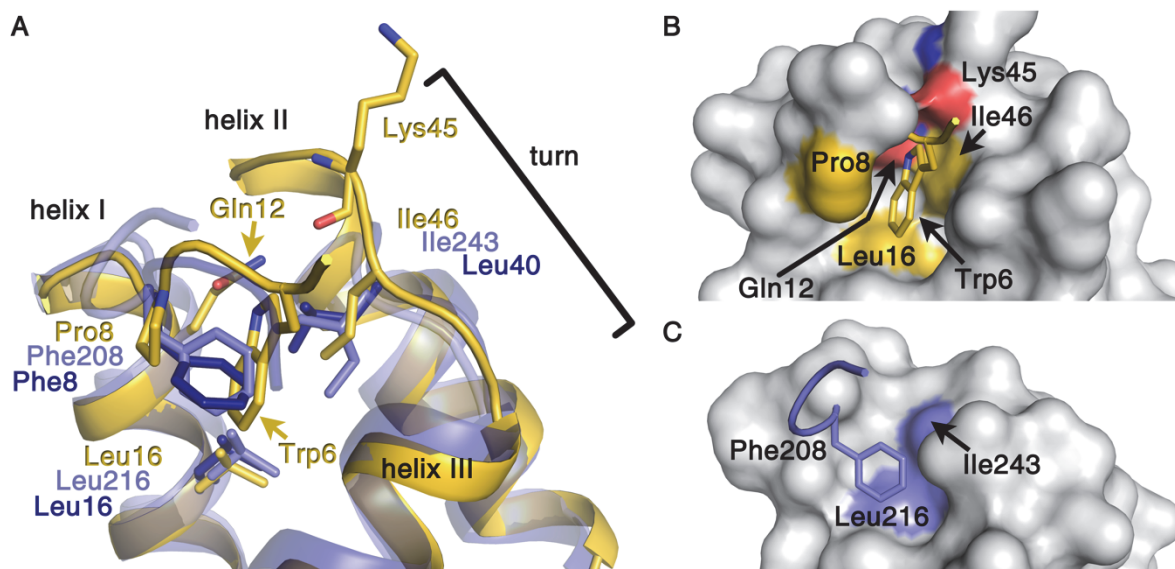


Fig. 14: Stabilization of the N-terminal arm in WUS HD compared to other published homeodomain structures. A) Superposed structures of WUS HD (yellow), EN HD (dark blue, transparent) and EXD HD (slate, transparent) displayed as ribbons; side chains of amino acid residues involved in the stabilization of the N-terminal arm are depicted as sticks; structures were cut N-terminal of homeodomain position 5 for reasons of clarity. B) Surface of WUS HD (grey) is shown for amino acid residues C-terminal of Trp6 (typ6); coloured surface illustrates the pocket enclosing the Trp6 (typ6) side chain formed by the side chains of Leu16 (typ16), Ile46 (typ40), Gln12 (typ12) and the main chain of Lys45. C) Surface of EXD HD (grey) is shown for amino acid residues C-terminal of Phe208 (typ8); slate coloured surface depicts the hydrophobic pocket formed by residues Leu216 (typ16) and Ile243 (typ40), which accommodates Phe208 (typ8).

amino acids at corresponding positions of the homeodomain and both bulky side chains are buried deeply within the structure (Table 9). However, a closer examination revealed major functional differences between the Phe236 (typ6) in HNF1 β HD and Trp6 (typ6) in WUS HD. Whilst Trp6 (typ6) itself links the N-terminal arm with the helix bundle, Phe236 (typ6) only contacts Trp238 (typ8), which in turn acts as the actual anchor. Resembling residue Phe/Tyr8, Trp238 (typ8) interacts with Leu246 (typ16) and Val291 (typ40) in HNF1 β HD (Fig. 16). Therefore Phe236 (typ6) appears to merely function as an additional stabilization of the N-terminal arm in HNF1 β HD on top of the usual and ubiquitous docking via the homeodomain position 8, which is in contrast to Trp6 (typ6) in WUS HD. The manner of stabilizing the N-terminal arm present in WUS HD displays several unique characteristics distinct from published homeodomain structures. Not only is a tryptophan residue or similar amino acid at position 6 of the homeodomain highly uncommon (Fig. 1), but also two of the amino acid residues, which form the pocket enclosing the Trp6 (typ6) side chain, Pro8 (typ8) and Lys45, are not available for this task in most other homeodomains. In general, the amino acid residue corresponding to Pro8 (typ8) is occupied with linking the N-terminal arm with the helix

bundle and thus may not be used for other purposes. Moreover Lys45 is part of DR II in WUS HD (see section 3.2.1) and thus simply not existent in other homeodomains. These aspects and the apparent differences to other published homeodomain structures imply that the docking of the N-terminal arm by Trp6 (typ6) can be considered a characteristic of WUS HD.

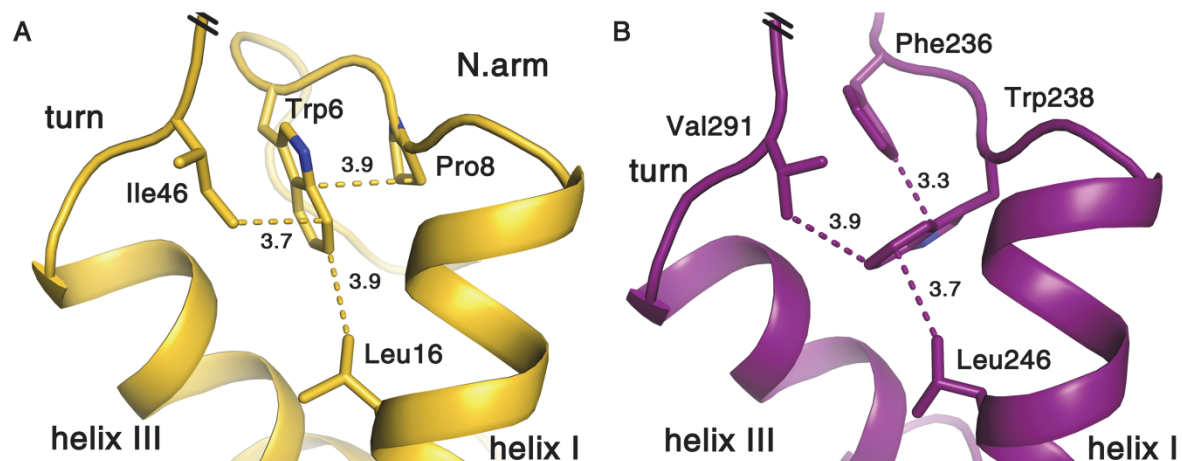


Fig. 15: Hydrophobic network established by Trp6 (typ6) in WUS HD (A) in comparison to Phe236 (typ6) in HNF1β HD (B). Structures are displayed as lines connecting the C α -atoms; for reasons of clarity helix II and most of the turn were removed, the cuts are indicated by two black parallel lines; side chains of amino acid residues participating in the hydrophobic network are shown as sticks; hydrophobic interactions are indicated as dashed lines; distances are given in [Å].

3.2.3.2 The characteristic conformation of DR I results from the integration of additional amino acids and a distinctive hydrophobic core amino acid residue

As stated in section 3.2.1 two extra residues are integrated in the loop region, which connects helix I and helix II. The residues Phe/Tyr20 and Pro/Leu26 are highly conserved in the entire homeodomain family (Fig. 1) and give some indication about the approximate identity of the additional amino acids of WUS HD in this region. The corresponding residues in WUS HD are Tyr20 (typ20) and Pro28 (typ26) (Fig. 16 E), respectively, and thus confirm the integration of two additional amino acids in between. The resulting divergent region (DR I), however, seems to comprise more than two residues. In case of superimposing WUS HD and the two typical homeodomains (Fig. 10 A and G) and HNF1β HD (Fig. 10 C and I) the computer program depicts Asn22 and Asn23 of WUS HD as gapped residues without structural equivalents and aligns the adjoining Ala24 of WUS HD to amino acid residue 22 of typical homeodomains. But when taking a closer look at the superposed structures Ala24 of WUS HD is located between residues 22 and 23 of typical homeodomains and does not seem to correspond properly to either one (Fig. 16 B). Comparison of the WUS HD structure and TALE homeodomains provides

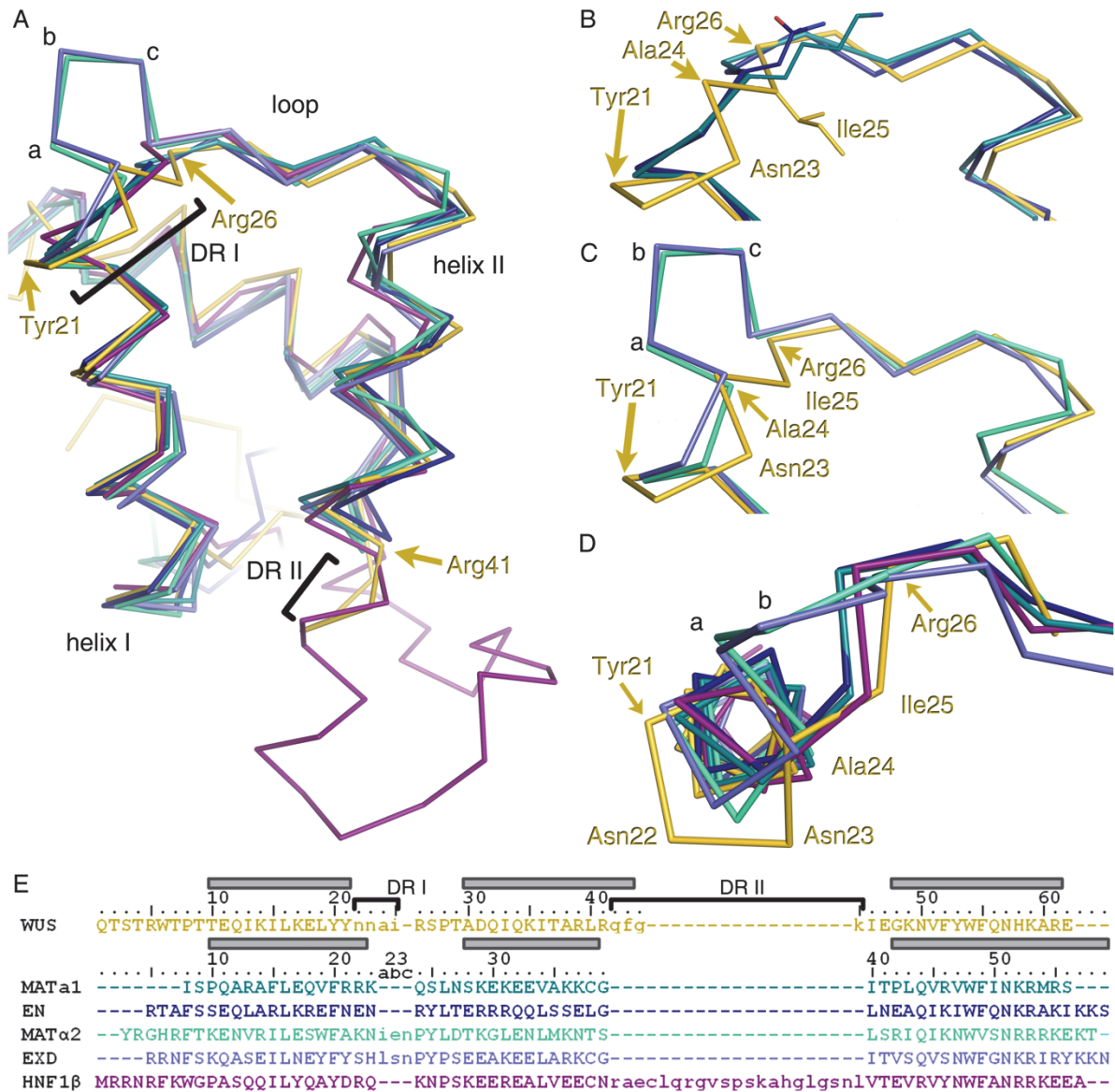


Fig. 16: Effects of additional amino acids on the structure of WUS HD and on other homeodomains. A-D) Superimposed structures of the homeodomains of WUS (yellow), MATA1 (teal), EN (dark blue), MATA2 (turquoise), EXD (slate), HNF1 β (purple) displayed as lines connecting the C α -atoms. For reasons of clarity parts of the structure were removed in B-D. A) Overview of the location of DR I and DR II of WUS HD. B) and C) Structural conformation of DR I compared to typical (B) and TALE (C) homeodomains. D) Top view on DR I. E) sequence alignment according to superimposed structures; boxes indicate the location of helices as described in previous publications (Fraenkel and Pabo 1998; Passner, Ryoo et al. 1999); brackets depict amino acids comprising DR I and II of WUS HD.

additional information. The TALE homeodomain positions 23 and 24, which border the additional amino acids “a, b and c” (Fig. 16 E), superimpose rather well with Ala24 and Arg26 of WUS HD (Fig. 16 C) suggesting that Ile25 and either Asn22 or Asn23 are the two amino acid residues of WUS HD without structural equivalent in other homeodomains.

Taken together, only Tyr21 and Arg26 of WUS HD appear to be reasonably well aligned to corresponding residues in all of the examined homeodomain types. But no distinct structural equivalent can be assigned to either of the amino acids in between. Therefore, though only two additional amino acids are integrated in this region, DR I comprises four amino acid residues ranging from Asn22 to Ile25 (Fig. 16A and E). These findings are in contrast to the assumption that additional amino acids of atypical homeodomains are simply looped out of the typical homeodomain structure (Bürglin 1994) and encouraged a more detailed analysis of the conformation of DR I. The bulging out of additional amino acids in the loop region, as is apparent in TALE homeodomains, for example, seems to be prevented in WUS HD by Ile25, which is not only part of DR I but also a distinctive amino acid residue of the hydrophobic core of WUS HD (section 3.2.2, Fig. 11). The side chain of Ile25 protrudes towards the centrepiece of the homeodomain, where it establishes an extensive network with amino acid residues of the hydrophobic core (Fig. 8). Thus in contrast to other homeodomains the loop of WUS HD is connected to the hydrophobic core not only by Pro28 (typ26), which corresponds to the conserved Pro/Leu26 in typical homeodomains (Fig. 1), but additionally by Ile25 rendering the loop of WUS HD highly rigid (Fig. 5, Table 6). As a consequence the pressure on the WUS HD structure created by the additional amino acids accumulates and manifests in a more densely packed loop, which is tilted towards the opposite direction of the loop of typical homeodomains (Fig. 16 B and C), and a dislocation of Asn22 and Asn23 extreme enough to cause a shortening of helix I (Fig. 16 D and E).

Taken together the integration of two additional amino acids results in the structural deviation of a region, named DR I, spanning four amino acids (Asn22 to Ile25). DR I adopts a conformation distinct from comparable regions in other published homeodomain structures (Fig. 16 A-D). A central player of this novel conformation is Ile25, a characteristic amino acid residue of WUS HD (see section 3.2.2) with unusually extensive contacts to the hydrophobic core (Figs. 8 and 16 B).

3.2.3.3 Simultaneous extension of helix II and the adjacent turn preserve the HTH-motif in WUS HD

Determination of the structure of the ANTENNAPEDIA homeodomain confirmed the hypothesis that helix II and III fold into a helix-turn-helix (HTH) motif (Qian, Billeter et al. 1989), which had been well known from prokaryotic DNA binding proteins, such as the viral CRO repressor (Anderson, Ohlendorf et al. 1981) or the LAC repressor of *E. coli* (Matthews,

Ohlendorf et al. 1982). However some atypical homeodomains harbour extra residues in this region perturbing the HTH motif. The 21 additional amino acids of HNF1 β HD for example result in the C-terminal elongation of helix II and the adjoining connecting region to helix III. Thus, the sharp turn, which typically connects the helices II and III, is replaced by an extended loop region and consequential the HTH motif is absent in HNF1 β HD (Fig. 17). Four of the six additional amino acids of WUS HD cause the deviation of the amino acid residues Gln42 to Lys45 (DR II) from that of the typical homeodomain conformation (see section 3.2.1). The amino acid residues of DR II can be aligned to corresponding amino acids in HNF1 β HD. But despite this partial overlap, the two regions differ in some aspects. Similarly to HNF1 β HD also helix II of WUS HD is extended at its C-terminus, but only by three residues (Fig. 16E). One

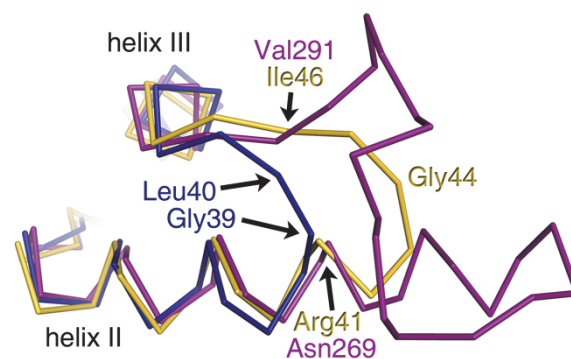


Fig. 17: DR II of WUS HD. Superimposed structures of WUS HD (yellow), EN HD (dark blue) and HNF1 β HD (purple) displayed as lines connecting the C α -atoms.

extra amino acid seems to be required to bridge the increased distance between the helices II and III (Fig. 17), facilitating the maintenance of the relative orientation of the helices to each other despite the elongation of helix II. Hence, in contrast to HNF1 β HD, the canonical HTH motif is preserved in WUS HD, with the variation of a turn comprising four amino acids instead of the usual three of typical homeodomains.

3.3 WUS HD represents a novel homeodomain subtype

With very few exceptions, only typical and TALE homeodomains of animal origin have been the subject of research so far and no structure of a plant homeodomain has been determined to date. Moreover the protein fold adopted by typical and TALE homeodomains is highly specific leading to a low diversity in the pool of determined homeodomain structures, which complicates predictions about the structural effects of additional amino acids. The determination and analysis of the *de novo* and thus unbiased structure of the WUSCHEL homeodomain commences to fill this gap.

WUS HD represents the first structure of a plant homeodomain. The comparative analysis of WUS HD to published animal and yeast homeodomains shows that its general structure, composed of a specific assembly of three α -helices, has been preserved during evolution (see section 3.2.1). Thus the structure of WUS HD provides further evidence that indeed all homeodomains share a common eukaryotic origin, which has been indicated by sequence similarity only so far (Mukherjee, Brocchieri et al. 2009).

Aside of the highly conserved regions, however, WUS HD differs in several aspects from that of published homeodomain structures. One is a novel manner of stabilizing the N-terminal arm. In most homeodomain structures, the highly conserved residue Phe/Tyr8 (Fig. 1) links the N-terminal arm with the helix bundle, whereas in WUS HD this task is carried out by the more N-terminal residue Trp6 (Fig. 9 and Fig. 14). Another difference is Divergent Region I (DR I), which comprises a characteristic conformation adopted by the N-terminal portion of the loop (Fig. 16). The third major difference between WUS HD and other homeodomain structures are modifications of the HTH-motif typically formed by the helices II and III and the connecting turn. In WUS HD the extension of helix II by three amino acids is balanced by the integration of one additional amino acid in the turn, which together constitute DR II (Fig. 16 A and D, Fig. 17). This way the HTH-motif and thus the relative orientation of helix III in WUS HD is preserved despite the integration of four additional amino acids in this region.

The relative orientation of the three helices seems to be essential for homeodomain function implied by the high degree of structural conservation throughout eukaryotic evolution. Only few regions of the homeodomain seem to be allowed slight modifications. Also in WUS HD the extra amino acids are inserted in regions comparable to the sites of integration apparent in other atypical homeodomains, e.g. the here compared structures of TALE homeodomains or HNF1 β HD. However, the conformations adopted by DR I and DR II in WUS HD do not resemble previously determined structures of atypical or typical homeodomains. These findings suggest that the X-ray analysis of WUS HD presented in this study illustrates a novel homeodomain subtype.

4 Discussion

4.1 The *wus-7* allele and putative structural basis for impaired function of *wus* mutant protein

In the course of studying the stem cell niche of *Arabidopsis thaliana* several mutations have been introduced into the *wuschel* gene that lead to varying degrees of impaired or even loss of function. One of the weaker alleles, *wus-7*, results in an attenuated *wus* activity in mutant plants (Graf, Dolzblasz et al. 2010), which is likely to relate to the observed decreased binding to target sites *in vitro* of the according mutant *wus* protein (Perales, Rodriguez et al. 2016). The *wus-7* mutation (G230>A) results in an exchange of glycine for glutamate at residue 77 of the full length WUSCHEL protein (wusG77E) (Graf, Dolzblasz et al. 2010), which is equivalent to Gly44 of the homeodomain of WUSCHEL. The structure of WUS HD illustrates that Gly44 is the first amino acid of the turn, which connects

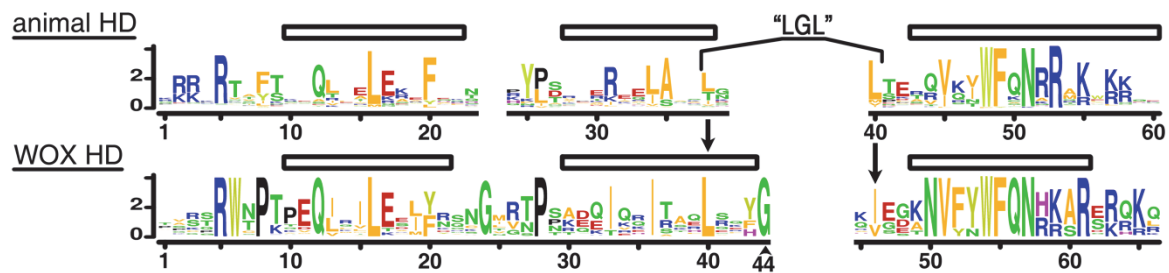


Fig. 18: Comparison of amino acid composition of animal and WOX homeodomains. Sequence logo of animal HD as in Fig. 1; sequence logo of WOX HD is summarized from Fig. SX; please note, that WOX homeodomains consist of either 65 or 66 amino acids leading to a gap at residue 23 inserted by the sequence alignment program in case of 65-amino acid WOX homeodomains; the alignment of the frequency maps is based on the superposed structures of WUS HD and other animal homeodomains according to Fig. 16; sequence logos are represented as bit scores (y-axis) in function of the homeodomain residue (x-axis); relative positions of the helices in WUS HD and typical animal homeodomains (Fraenkel and Pabo 1998; Passner, Ryoo et al. 1999) are indicated by black rectangles above the sequence logos.

helix II and the DNA recognition helix of WUS HD (Fig. 4A, Fig. 18). At structural equivalent positions of HTH motifs in prokaryotic repressor proteins a glycine seems mandatory most likely to avoid steric interference (Brennan and Matthews 1989). Though the stereochemical requirements for the first residue of the turn appear to be relaxed in homeodomain proteins (Gehring, Affolter et al. 1994), glycine is still amongst the most frequent amino acids at this position, e.g. Gly39 of EN HD in Fig. 17, which forms together with the two neighbouring residues the well known “LGL” motif at positions 38 to 40 (Fig. 18). Inferring from its strict

conservation Gly44 seems essential in the family of WOX homeodomains (Fig. 18), which might be correlated with the four additional amino acids in this region and the resulting extension of helix II (DRII, Fig. 17). Thus the exchange of glutamate in place of glycine may violate sterical constraints inflicted by the turn, and the product of *wus-7* is likely to display impaired DNA binding, which is consistent with the observed lowered affinity to target sequences (Perales, Rodriguez et al. 2016) and the attenuated *wus* activity in *wus-7* plants (Graf, Dolzblasz et al. 2010)

Furthermore it has been suggested that residue Gly44 of the WUSCHEL homeodomain is required for homodimerization (Perales, Rodriguez et al. 2016). However, in the course of the experimental phase of this study, indications of a possible dimerization of the WUSCHEL homeodomain were observed only in the presence of DNA, e.g. analysis of DNA binding with ITC, but never during the process of protein purification or in the crystal structure (data not shown). This is consistent with the data shown in the publications on dimerization of WUSCHEL (Perales, Rodriguez et al. 2016; Rodriguez, Perales et al. 2016). These findings indicate that homodimerization of the WUSCHEL homeodomain is likely to be DNA-mediated and may not depend predominantly on the direct interaction of the two homeodomains via Gly44. A lowered DNA binding affinity is likely to also affect DNA-mediated interaction with co-factors and thus may explain the reported decrease in homodimerization of *wus* proteins exhibiting the mutation G44E in their homeodomain.

4.2 Possible effects of novel characteristics identified in WUS HD on DNA recognition

4.2.1 Generation of models illustrating the potential DNA recognition of WUS HD

A fast amount of structural, biochemical, mutational and functional analyses on homeodomain DNA recognition is available. In combination with the high degree of conservation of the homeodomain structure as well as its manner of binding to DNA all these results may be applied to other homeodomains (section 1.1.2). The structure of WUS HD determined and analysed in this study facilitates the certain identification of WUS HD residues corresponding to common homeodomain specificity determinants (section 1.1.3). Curiously many of the novel characteristics of WUS HD were found to be imbedded in regions of the homeodomain known to be directly involved in DNA recognition and determination of sequence specificity, for example the N-terminal arm and the loop as well as specific residues of the recognition

helix (section 1.1.3). The stabilization of the N-terminal arm and the conformation of the loop are distinctive characteristics of WUS HD and have not been described elsewhere, yet (section 3.2.3). Also the residues Phe52 and Tyr53 of helix III in WUS HD, which correspond to the specificity determinants at the positions 46 and 47 in typical HD (Fig. 3), display highly unusual amino acids (Fig. 18). In addition implications on the potential effects of the structural characteristics determined for WUS HD on DNA recognition may also add to understanding the structural evolution of other atypical homeodomains.

Thus it seemed worthwhile to examine the putative effects of these unusual features of WUS HD on its DNA binding properties in more detail. In order to better address these questions it seemed inevitable to model a complex of WUS HD bound to DNA. The in the following proposed and discussed model makes no claim on absolute accuracy, rather with regard to the high degree of conservation of homeodomain DNA recognition (section 1.1.2) and in the absence of a true complex of WUS HD bound to DNA, the model aims to illustrate the possibilities and to serve as a good basis for new hypotheses and subsequent research.

Comparisons of monomeric homeodomain structures with others that were solved in complex with DNA imply that the main chain of a homeodomain does not undergo major structural changes upon DNA binding (Clarke, Kissinger et al. 1994; Fraenkel, Rould et al. 1998). This also seems to be reflected in the high similarity of WUS HD to other homeodomains regardless of whether the other structure was solved as monomer or in complex with DNA (section 3.2.1, Table 8). Additionally the invariant Asn51 of the WFXN motif establishes a distinct bidentate contact with an adenine at position 3 (A3) (Fig. 3B, Table 2). This interaction has been detected in all homeodomain-DNA complexes (section 1.1.2), which predestines it as landmark for the correct positioning of the recognition helix of WUS HD in the major groove. Consequential the model is likely to depict interaction of helix III with the DNA quite well. The conformation of the N-terminal arm and its specification of nucleotide bases, however, strongly depend on its interaction with the DNA backbone (section 1.1.3.2). Thus in case of the N-terminal arm the model is to be interpreted with more care.

For the model WUS HD was aligned with multiple independently determined homeodomain structures in complex with DNA, which supposedly allows an estimation of the variation in DNA binding amongst different homeodomains and thus may facilitate a better interpretation of the model. As reference structures serve co-crystals with DNA of the homeodomains of MATa1 (teal, PDB ID 1YRN), EN (dark blue, PDB ID 3HDD), MATa2 (green cyan, PDB ID 1APL), EXD (light blue, PDB ID 1B8I) and HNF1 β (purple, PDB ID 3H8R), all of

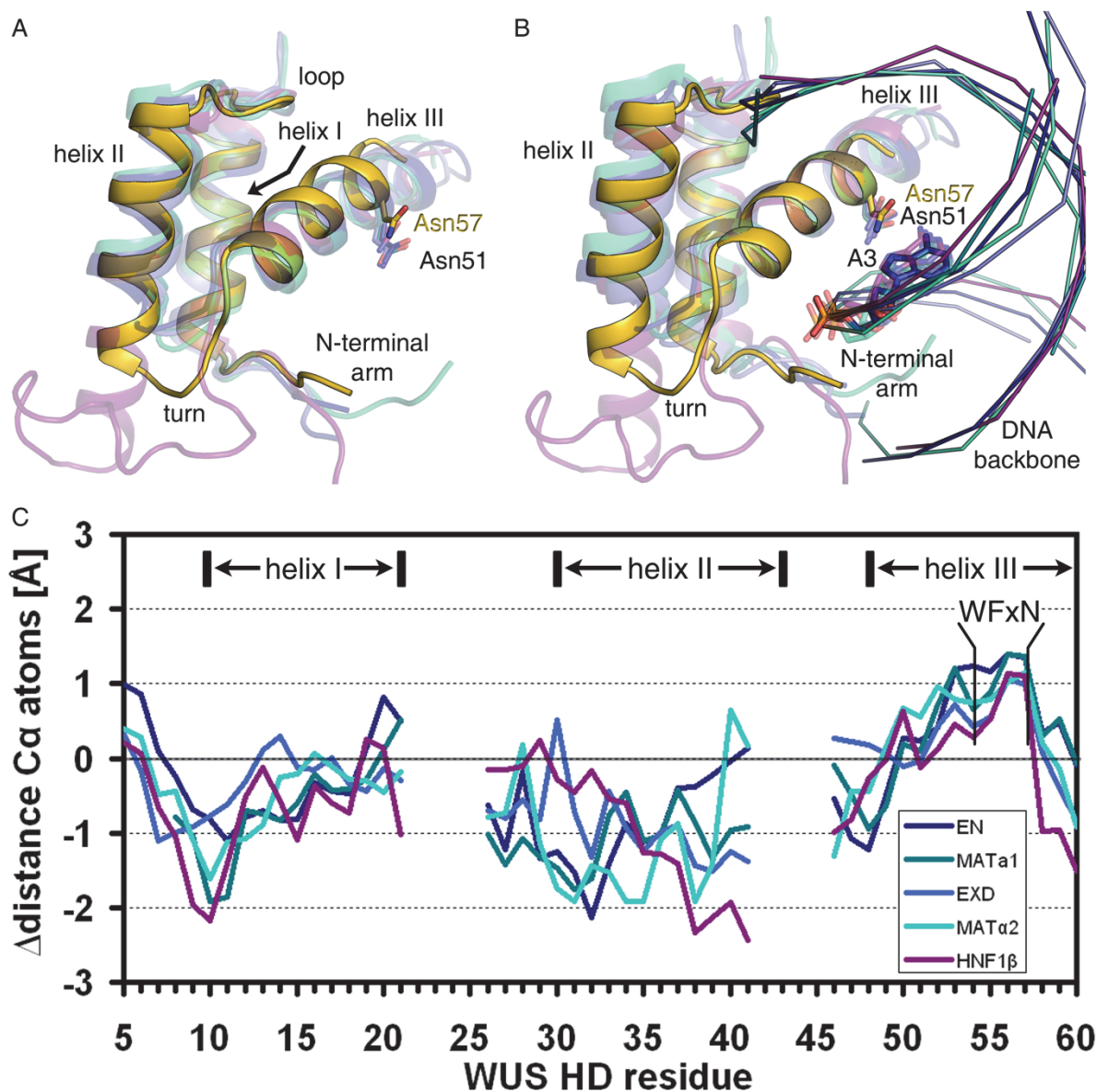


Fig. 19: Differences in the accuracy of the superimposition between two kinds of structural alignments. Homeodomains are depicted as ribbons, the DNA backbone as lines connecting the phosphate groups; side chains of the asparagine of the WFXN-motif and A3 of the DNA core sequence are displayed as sticks. A) Alignment to chB of WUS HD (yellow) with best global fit, which was also used for the structural analysis in section 3.2. B) Alignment to chB of WUS HD (yellow) with emphasis on the WFXN-motif. C) Distance [Å] between Cα atoms of chB of WUS HD and corresponding residues in reference structures was determined after global fit alignment (A) and alignment emphasised on WFXN-motif (B) and the difference (Δ) calculated; positive Δ distance values relate to an enhanced superposition in (B) compared to (A); the positions of the helices in WUS HD are indicated by vertical lines above the graphs.

which were also utilized for structural comparison with WUS HD in this study (section 3.2). Only in the case of EN HD another structure was selected, since the one used previously (PDB ID 1ENH) was solved without DNA. The kink at the C-terminus of helix III apparent in the the WUS HD copy chA seems to be caused by crystal packing and is unlikely to

represent a native conformation. Thus putative DNA recognition of WUS HD is analysed mainly by means of chB. Structural alignment was performed by superimposing the homeodomain motifs in helix III (WFXN) that include the aforementioned invariant Asn51 as well as the highly conserved residues Trp48 and Phe49, which dock helix III to the hydrophobic core (sections 3.1.4 and 3.2.2) (reviewed in Bürglin 1994). For the model of WUS HD bound to DNA this sort of alignment was favoured over the usual global fitting in order to focus on the correct insertion of helix III into the major groove and to mimic DNA recognition (compare Figs. 19A and B). To determine regions with enhanced or diminished superimposition, distances between C α atoms of corresponding residues were determined for both kinds of structural alignments and the values subtracted. A decrease in accuracy of superimposition was mainly detected for helix II and residue 10 (Fig. 19 C). The alignment of helix I and corresponding residues of the loop seems to be only mildly affected. In addition to the intended enhancement of the superimposition of helix III also the alignment of residue 5 seems to have improved slightly (Fig. 19 C), indicating a possible structural linkage of the two major regions of the homeodomain responsible for DNA sequence specificity (see section 1.1.3.1 and 1.1.3.2, Fig. 3). Thus the putative position of helix II relative to the DNA as is suggested in the model of the WUS HD-DNA complex is to be interpreted with care. But in case of the other homeodomain regions the loss in accuracy of the superimposition due to the focus on the WFXN-motif seems negligible. So the model may serve as a good basis for discussing the putative DNA recognition of WUS HD.

4.2.2 Residues of the putative DNA interaction surface of WUS HD

According to covariation analyses as well as structural and mutational data on DNA binding of homeodomains, the interaction surface is composed of residues of the N-terminal arm (2, 3, 5, 6 and 8), the loop (25), the N-terminus of helix II (28 and 31) and nearly all of helix III (43, 44, 46, 47, 48, 50, 51, 53, 54, 55 and 56) (numbering is according to the position in typical homeodomains; Fig. 3, Table 2; for more details see section 1.1.3). In WUS HD these positions correspond to Thr2, Ser3, Arg5, Trp6 and Pro8 of the N-terminal arm, Ser27, which is located in the loop, Ala30 and Ile33 at the N-terminus of helix II and the following residues of helix III: Lys49, Gln50, Phe52, Tyr53, Trp54, Gln56, Asn57, Lys59, Ala60, Arg61 and Glu62 (Fig. 16). In order to facilitate easy comparison with other publications the position of the structural equivalent in typical homeodomains will be referred to in parentheses after the respective WUS HD residue. The position of residues Thr2 (typ2), Ser3 (typ3), Arg61 (typ55),

Glu62 (typ56) could not be determined for chB of the crystal structure of WUS HD. In case of the N-terminal arm only basic amino acids seem to possess the biochemical properties required for interaction with DNA bases ((Noyes, Christensen et al. 2008; Dror, Zhou et al. 2014), for more details see section 1.1.3.2), suggesting that neither Thr2 nor Ser3 are likely to be involved in DNA recognition of WUS HD. In other homeodomains Arg55, the residue corresponding to Arg61 (typ55) of WUS HD, was shown to correlate (Noyes, Christensen et al. 2008) and interact (Table 2) with a guanine at position 2 of the DNA target sequence (for more details see section 1.1.3.1). The other residues, whose position is resolved in chB, constitute the putative

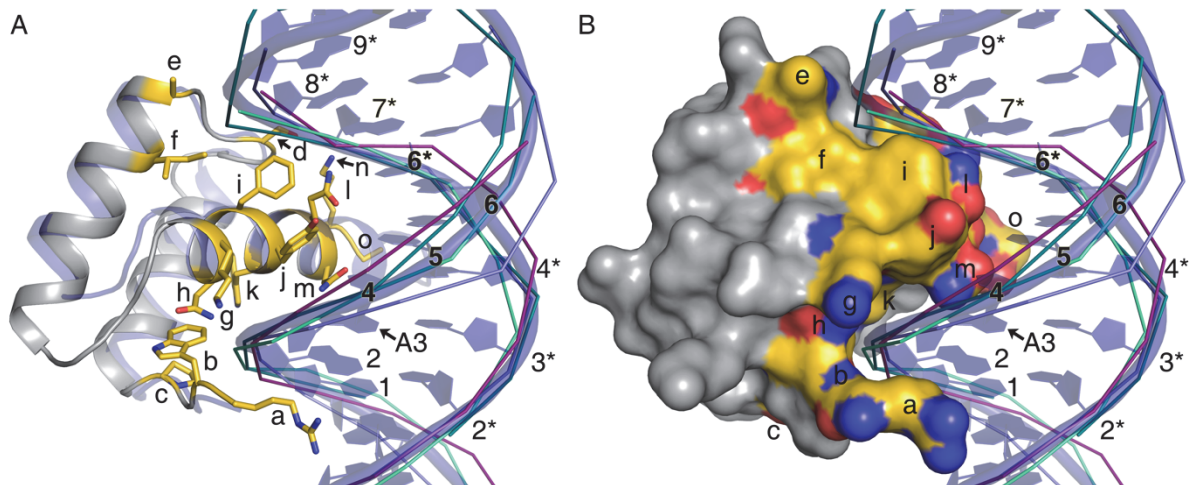


Fig. 20: Putative DNA interaction surface of WUS HD. A and B) Superimposition of homeodomains solved in complex with DNA and WUS HD as described in section 4.2.1 and shown in Fig. 22 B; the DNA of the complex with EN HD is shown with nucleotides (blue, transparent), the DNA helices of the other structures are indicated by lines connecting the phosphate groups; labelling of DNA nucleotides is as described in section 1.1.2; WUS HD is coloured grey, only residues of WUS HD that correspond to amino acids of the conserved DNA interaction surface (Fig. 3) are highlighted in yellow: Arg5 (a), Trp6 (b), Pro8 (c), Ser27 (d), Ala30 (e), Ile33 (f), Lys49 (g), Gln50 (h), Phe52 (i), Tyr53 (j), Trp54 (k), Gln56 (l), Asn57 (m), Lys59 (n), Ala60 (o). A) EN HD (blue, transparent) superposed with WUS HD (grey and yellow) displayed as ribbons; side chains of residues composing the putative DNA interaction surface of WUS HD are shown as sticks. B) Surface of WUS HD.

DNA interaction surface of WUS HD in the model (Fig. 20) similarly to other known homeodomain-DNA complexes (Fig. 3). The DNA backbone may be contacted by the main chain of Trp6 (typ6) (“b” in Fig. 20) and the side chains of residues Ser27 (typ25) of the loop (“d” in Fig. 20) and Lys49 (typ43), Gln50 (typ44), Phe52 (typ46), Trp54 (typ48), Lys59 (typ53) of helix III (“g”, “h”, “i”, “k” and “n” in Fig. 20). The side chains of Arg5 (typ5) and Tyr53 (typ47), Gln56 (typ50), Asn57 (typ51), Ala60 (typ54) protrude into the minor and major groove of the DNA in the model (“a”, “j”, “l”, “m”, “o” in Fig. 20), where they may interact with DNA bases and thus determine the sequence specificity of WUS HD. Therefore the

putative DNA interaction surface of WUS HD largely overlaps with findings of previous publications on DNA binding by homeodomains (Fig. 3). Only three residues, which usually contact the DNA backbone, seem to be excluded from the DNA interaction surface of WUS HD: Pro8 (typ8), Ala30 (typ28) and Ile33 (typ31) (“c”, “e”, and “f” in Fig. 20). In case of Pro8 (typ8), this is likely to relate to the novel way of linking the N-terminal arm with the helix bundle in WUS HD, which utilizes Trp6 (typ6) rather than Phe/Tyr8. In WUS HD Trp6 (typ6) projects into the space usually occupied by the side chain of residue 8 (Fig. 14) and hence may block Pro8 (typ8) from participating in DNA recognition. In case of Ala30 (typ28) and Ile33 (typ31) the distance between their side chains and any part of the DNA backbone is too big to suggest a potential role in DNA recognition (Fig. 20). This seems particularly interesting in case of Ile33 (typ31). It replaces a highly conserved arginine (Arg31) (Fig. 18) that commonly interacts with a phosphate group of the DNA backbone (Fig. 3). Moreover the extreme conservation of Ile33 in WOX homeodomains (Fig. 18) suggests a preserved function in this homeodomain family. However neither the structure of WUS HD nor the model of WUS HD bound to DNA indicate any structural constraints for Ile33 or a potential function in DNA recognition. Therefore possible explanations for the evolutionary constraints that seem to work on this residue need to await a more detailed examination in subsequent studies.

Summarized the model of WUS HD bound to DNA suggests that its interaction surface is composed of similar residues as has been published for other homeodomains. Only three contacts to the DNA backbone may be missing in case of WUS HD, which are commonly established by residues 8, 28 and 31 and which correspond to Pro8, Ala30 and Ile33 of WUS HD.

4.2.3 Uncommon amino acids of helix III may affect sequence recognition of WUS HD

In WUS HD two residues of the recognition helix, which are highly likely to be part of the DNA interface, display quite uncommon amino acids, Phe52 (typ46) and Tyr53 (typ47) (Fig. 20 and Fig. 18). The structural equivalents of Phe52 (typ46) often are basic amino acids (Fig. 18) that contact the DNA backbone or - though more rarely - establish a water mediated contact to nucleotide bases (Fig. 3A, see sections 1.1.3.1 and 1.1.3.2). In the model of WUS HD bound to DNA the side chain of Phe52 (typ46) is placed in close proximity to the sugar moiety of nucleotide 8* (Fig. 21), which is consistent with previous reports on its general function, the interaction with the DNA backbone. But in addition the phenyl ring of Phe52 (typ46) is

positioned in immediate vicinity of the bases of nucleotides 7* (Fig. 21) indicating the potential of Phe52 (typ46) for direct base specification. Also the tyrosine at position 53 (typ47) of WUS HD seems peculiar at first glance. The two most frequent amino acids at corresponding positions are isoleucine and valine (Ile/Val47, Fig. 18), which both determine a thymine at position 4 of the core DNA sequence (Table 2, see section 1.1.3.1). In the model for DNA recognition of WUS HD the position of the hydrophobic portion of Tyr53 (typ47) matches that of Ile/Val47 exemplified by Ile47 of EN HD in Fig. 21. In contrast to isoleucine or valine, however, the bulky side chain of tyrosine is likely to reach much deeper into the major groove. In the model the terminal hydroxyl group of Tyr53 (typ47) directs towards the nucleotide bases at position 5 of the DNA helices (Fig. 21). Thus Tyr53 (typ47) of WUS HD may be a sequence determinant for position 4 similarly to Ile47/Val47 of other homeodomains and in addition may influence the identity of the neighbouring nucleotide at position 5 of the DNA binding sequence.

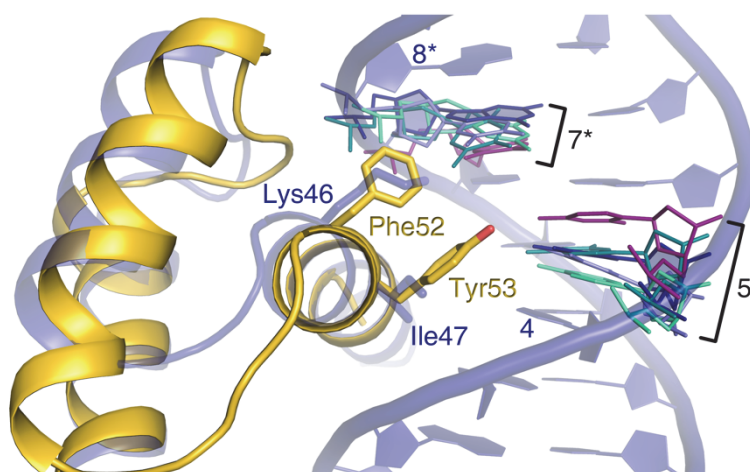


Fig. 21: Potential DNA recognition by Phe52 and Tyr53 of WUS HD. Structure of WUS HD (yellow) superimposed with EN HD in complex with DNA (blue, transparent) are displayed as ribbon; residues Phe52 and Tyr53 of WUS HD and corresponding residues in EN HD are depicted as sticks; all nucleotides at positions 5 and 8* of the DNA helices in complex with the analysed homeodomains are shown as sticks.

4.2.4 The differences in linking the N-terminal arm with the helix bundle apparent in WUS HD may affect its DNA recognition

Sequence specificity of the N-terminal arm are largely determined by its structural conformation conveyed by residues 4, 6 and 7, which position the base contacting residues, most commonly argenines at residues 2, 3 and 5, over the minor groove (Ekker, Jackson et al. 1994; Damante, Pellizzari et al. 1996; Joshi, Passner et al. 2007; Dror, Zhou et al. 2014). In the case of WUS HD the conformation of the N-terminal arm is greatly affected by Trp6 (typ6), which links the

N-terminal arm with the helix bundle in a novel way (section 3.2.3.2). Immediately adjacent to Trp6 (typ6) lies Arg5 (typ5), a highly conserved base contacting residue of homeodomains (Fig. 18, Fig. 3,) (Gehring, Qian et al. 1994; Passner, Ryoo et al. 1999; Chaney, Clark-Baldwin et al. 2005; Zhang, Larsen et al. 2011). Hence the structural changes in docking the N-terminal arm of WUS HD are likely to also result in differences in DNA recognition.

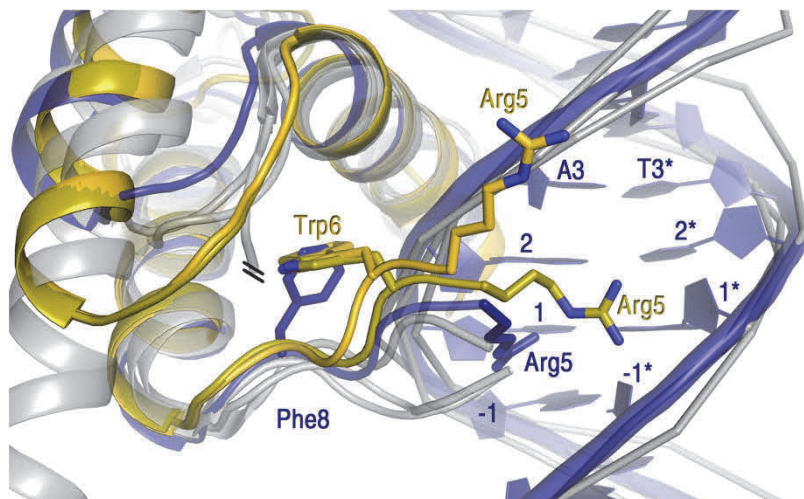


Fig. 22: Possible effects of Trp6 docking on DNA recognition of the N-terminal arm. Superposed structures of the two WUS HD copies, chA (yellow) and chB (olive), and four homeodomain-DNA complexes: EN HD (dark blue, PDB ID 3HDD), MAT α 2 HD (grey, PDB ID 1APL), EXD HD (grey, PDB ID 1B8I) and HNF1 β HD (grey, PDB ID 3H8R); homeodomains are depicted as ribbon, the DNA helices as lines connecting the phosphate groups of the DNA backbone, only the DNA helix of EN HD (blue) is illustrated as ribbon with nucleotides; the region N-terminal of residue 5 is not displayed for any structure; the side chains of Trp6 and Arg5 of both WUS HD copies and of Phe8 and Arg5 of EN HD are shown as sticks; region connecting helix II and III was cut in case of HNF1 β HD indicated by two parallel lines.

As is mentioned in section 3.1.2 the positions of the Trp6 (typ6) side chains in the two WUS HD copies (chA and chB) overlap nearly completely, though the locations of their C α atoms diverge by 1.2 Å (Fig. 5A and C), which seems to be the result of the tight grip of the residues enclosing the Trp6 side chain (Fig. 9). This finding indicates that the two conformations of the N-terminal arm displayed in chA and chB of the WUS HD crystal may already illustrate a great fraction of the possible maximum range of motion and larger movements may require the undocking of Trp6 (typ6) from its pocket. Thus even though the eventual position of the N-terminal arm can only be determined in complex with DNA, discussing the conformation of the N-terminal arm in chA and chB may already give a hint on the possible effects of Trp6 docking on the DNA recognition.

The most striking effect of Trp6 is that the N-terminal arm of WUS HD seems to be pulled closer to the helix bundle, in particular closer to helix III, when compared to homeodomains with Phe/Tyr8 fixation (Fig. 22). The model for DNA recognition of WUS HD indicates that Arg5 (typ5) may be positioned closer to nucleotide 2 of the DNA binding sequence (Fig. 22) in contrast to the corresponding amino acid residue in other homeodomains, which typically interacts with nucleotides 1 and/or -1* (Table 2, Fig. 22). Hence it seems likely that DNA recognition of the N-terminal arm in WUS HD is affected by its docking to the helix bundle via Trp6 (typ6).

4.2.5 The unique conformation of DR I of WUS HD may result in alterations of DNA backbone contacts

Animal homeodomains possess a highly conserved amino acid in the loop, Tyr25 (Fig. 18), which contacts the DNA backbone in determined structures of homeodomains bound to DNA (Fig. 3A and C). Furthermore co-variation analysis of mouse homeodomains and preferred DNA sequences identified residue 25 as important for DNA recognition (Berger, Badis et al. 2008) and mutation of Tyr25 was shown to be detrimental for homeodomain function in many cases (Furukubo-Tokunaga, Muller et al. 1992; Mathias, Zhong et al. 2001; Sato, Simon et al. 2004) (for more details see section 1.1.3.3). A possible structural basis for the severe effects of Tyr25 mutation may be given by MATa1 HD, in which a native Ser25 is unable to contact the DNA backbone (Fig. 3A) (Li, Stark et al. 1995). Like MATa1 HD also WUS HD displays a serine at the corresponding homeodomain position, Ser27 (typ25) (Fig. 16E). In contrast to Ser25 of MATa1 HD, however, Ser27 (typ25) of WUS HD is imbedded into a novel loop conformation distinct from typical as well as TALE homeodomains (see section 3.2.3.2). To further investigate the novel loop conformation and its potential effects on recognition of the DNA backbone, this region was analysed in more detail in the compilation of homeodomain-DNA complexes superposed with WUS HD as described in section 4.2.1.

At first glance the locations of the loop seem to differ quite a bit between the individual homeodomains (Fig. 23). However, when taking a closer look the ways in which Tyr25 interacts with the DNA are highly similar, exemplified by EN HD and MAT α 2 HD in Fig. 23B. Distances between the oxide atoms of the terminal hydroxyl group of Tyr25 and the phosphate group of nucleotide 6* range from 2.5 Å to 2.8 Å in the structures of EN HD (PDB ID 3HDD), EXD HD (PDB ID 1B8I) and MAT α 2 HD (PDB ID 1APL). With distances of either 4.4 Å or 4.3 Å the variation is even lower for the span between the phenyl rings of Tyr25 (CE1) and the

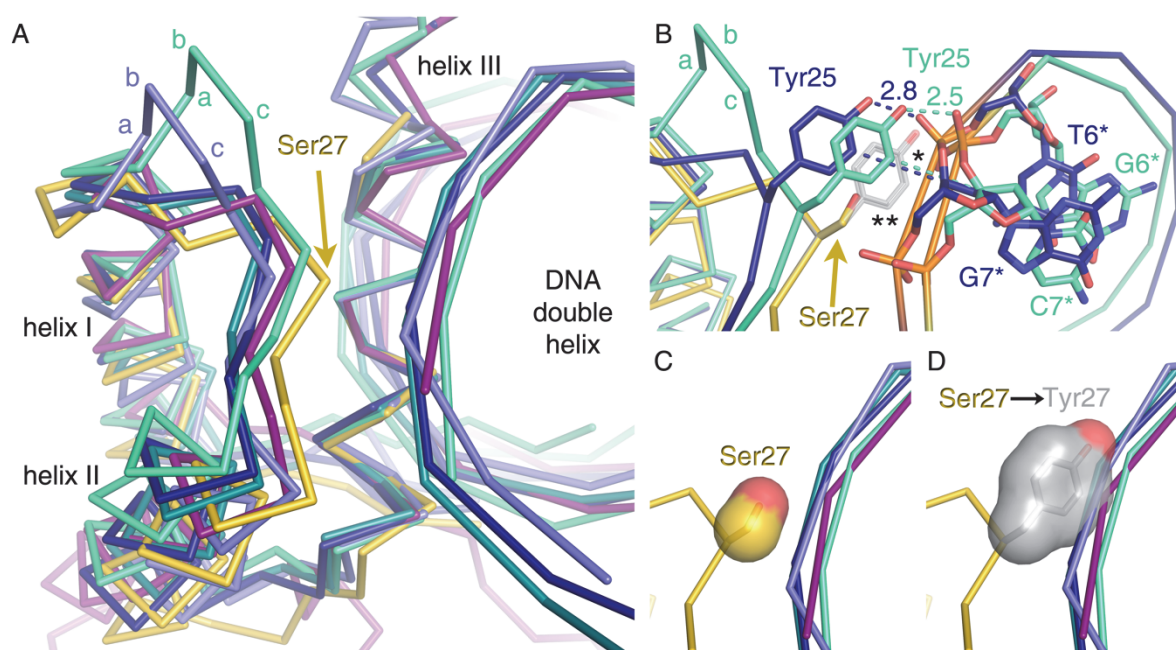


Fig. 23: Putative co-dependency of DR I and the lack of the conserved tyrosine at the DNA backbone contacting residue in the loop of WUS HD. A), B), C) and D) Top view on loop region of superimposed structures of homeodomain-DNA complexes and WUS HD chB (yellow); structural alignment was performed with emphasis on the WFXN motif as described in section 4.2.1; EN HD (dark blue, PDB ID 3HDD), EXD HD (slate, PDB ID 1B8I), MATa1 HD (teal, PDB ID 1YRN), MATa2 HD (green cyan, PDB ID 1APL), HNF1 β HD (purple, PDB ID 2H8R). A) Overview of all aligned structures; the C-terminal region of the loop of WUS HD is positioned closest to the backbones of the DNA helices. B) Interaction of Tyr25 with the DNA backbone is highly conserved (dashed lines); distances are given in [Å]; distance between CE of Tyr25 and sugar moiety of DNA backbone is 4.4 Å and 4.3 Å in case of EN HD and MATa2 HD, respectively (*). B and C) The side chain of Ser27 (typ25) fits into the model for DNA recognition of WUS HD; the side chain of Ser27 (typ25) is shown as yellow stick (B) and with surface in (C). B and D) In contrast a tyrosine at residue 27 (typ25) may interfere with DNA binding of WUS HD; putative side chain conformation of Tyr27 (typ25) in WUS HD is shown as grey stick in (B) (**) and with surface in (D).

sugar moieties of nucleotide 7*. Thus it appears that interaction of Tyr25 with the DNA is well conserved in homeodomains.

Helix I of WUS HD seems reasonably well aligned to the respective helices of the other homeodomain structures also when the superimposition is performed with emphasis on the WFXN motif in helix III (Fig. 19). In this sort of alignment structural divergencies between WUS HD and the aligned homeodomains increase strongly after DR I. As a consequence of DR I the whole loop region appears to be pushed closer towards the DNA backbones of the aligned homeodomain-DNA complexes (Fig. 23A and B). Also the side chain of Ser27 (typ25) of WUS HD is positioned in the immediate vicinity of the DNA backbones despite its very short reach (Fig. 23B and C). In comparison a tyrosine at residue 27 (typ25) of WUS HD may even interfere with DNA binding (** in Fig. 23B and D). Hence the model for DNA recognition of

WUS HD indicates that the lack of the conserved amino acid at the backbone contacting residue of the loop may be linked with DR I of WUS HD. However, the confirmation of the two hypotheses – impairment of WUS HD DNA binding by mutating Ser27 (typ25) to tyrosine and a possible linkage of serine at residue 27 (typ25) of WUS HD with DR I - require further experiments.

4.3 Implications on the evolution of the homeodomain structure and function in the WOX family inferred from the WUS HD structure and conservation of key amino acid residues

The structurally important Gln12 and the amino acids, which form the inner hydrophobic core of WUS HD (Leu16, Leu19, Pro28, Ile36, Leu40, Ile46, Val51 and Phe55) are highly conserved in the entire WOX family (Fig. 18 and Fig. SX) suggesting a similar overall structure of all WOX homeodomains. Furthermore Trp6 is invariant in the WOX family and thus docking of the N-terminal arm as described for WUS HD in section 3.1.5 is likely to be a characteristic of all WOX homeodomains as well. Also the modified HTH-motif is likely to be a common feature of WOX homeodomains indicated by the integration of four additional amino acids in the entire WOX family similarly to DR II of WUS HD (Fig. 24 and Fig. SX) and the high conservation of residue Gly44, which borders on the C-terminal end of helix II of WUS HD (section 4.1).

However some differences are noticeable that may affect structure and function of individual homeodomains of the WOX family. First a salt bridge established between residues Gln17/Arg52 (Baird-Titus, Clark-Baldwin et al. 2006), which is common to animal homeodomains, is not apparent in WUS HD, where the corresponding amino acid residues are Lys17 (typ17) and His58 (typ52) (see section 3.1.3). In WOX13- and WOX9-clade homeodomains, however, amino acids Glu17 (typ17) and Arg58 (typ52) are present at the equivalent positions (Figs. 27 and SX). Thus in WOX13- and WOX9-clade homeodomains the according connection may be apparent, but seems to have been lost or replaced in sequential steps in WUS-clade homeodomains. Experiments conducted with EN HD suggest a link between DNA binding properties and the salt bridge established by amino acids at the homeodomain positions 17 and 52, though none of them face the DNA contact surface. Like WUS HD also EN HD displays amino acids at the corresponding residues, which are unable to interact: Lys17 and Lys52. Curiously the combinations Glu17/Lys52 and Lys17/Glu52 seem to be structurally favoured by the EN HD fold over the native Lys17/Lys52 (Stollar, Mayor et al. 2003; Sato,

Simon et al. 2004). Both mutant proteins, however, show a substantial loss in target affinity suggesting a reciprocal relationship between homeodomain stability and its ability to strongly bind DNA. In case of the WOX family these findings indicate that the fold of WOX13- and WOX9-clade homeodomains may be stabilized by the Glu17/Arg58 salt bridge, whereas the focus could be shifted to DNA affinity in case of the homeodomains of the WUS-clade.

The other obvious difference between the WUS HD structure and that of most other WOX homeodomains is an additional amino acid in the region of DR I. The homeodomain of all WOX13-, WOX9-clade and also of most WUS-clade homeodomains comprises 65 amino acids. Only some *WOX* genes of the WUS-clade translate into a homeodomain of 66 amino acids, present most dominantly in angiosperm homologues of *WUS*, which includes the structure analysed in this thesis, but also *CaWUL*, the sole WUS-clade representative in the

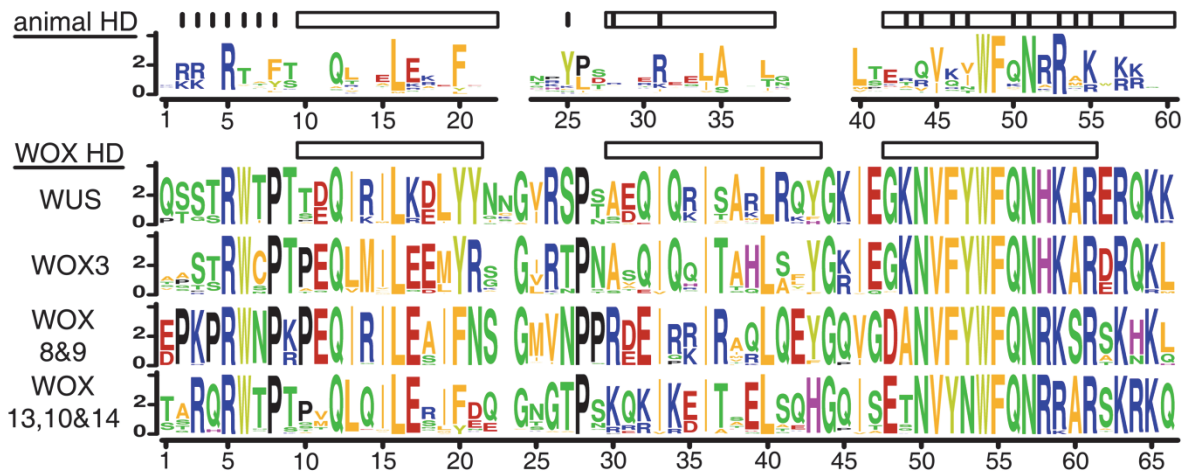


Fig. 24: Frequency of amino acids at corresponding positions in animal homeodomains and representatives of different WOX-clades. Sequence alignment is based on superimposed structures of WUS HD and other animal homeodomains as described in section 3.3; sequence logos are represented as bit scores (y-axis) in function of the homeodomain residue (x-axis); relative positions of helices in typical animal homeodomains (Fraenkel and Pabo 1998; Passner, Ryoo et al. 1999) and WUS HD are indicated by black boxes above sequence logos; as black vertical lines above the top logo depict sequence specificity determinants identified in animal homeodomains (Furukubo-Tokunaga, Muller et al. 1992; Ekker, Jackson et al. 1994; Damante, Pellizzari et al. 1996; Passner, Ryoo et al. 1999; Berger, Badis et al. 2008; Noyes, Christensen et al. 2008; Dror, Zhou et al. 2014);

fern *Ceratoperis richardii* (Nardmann, Reisewitz et al. 2009; Nardmann and Werr 2012). As mentioned in section 3.2.3.2 the conserved homeodomain residues Phe/Tyr20 and Leu/Pro26, which correspond to Tyr20 (typ20) and Pro28 (typ26) of WUS HD, set the boundaries for the potential sites of integration of the two additional amino acids of WUS HD. But only one of those two extra residues is found in all WOX homeodomains, whereas the other is distinct to WOX homeodomains comprising 66-amino acids. Aside of Phe/Tyr20 the compilation of WOX

homeodomain sequences reveals another highly conserved residue in this family, Gly24 (Fig. 24 and Fig. SX), which corresponds to Ala24 of the WUS HD structure (Fig. 4A). Thus the additional amino acid, which distinguishes 66- from 65-amino acid WOX homeodomains, is likely to be either one of the residues 21, 22 or 23. Moreover, homology modelling of 65-amino acid WOX homeodomains (described in section 2.18) indicates that amino acid residues until position 23 describe an α -helix (data not shown) as has been described in case of other homeodomain structures (Fig. 24). However, this suggests an elongation of helix I in 65-amino acid WOX homeodomains in contrast to the structure of the 66-amino acid homeodomain of WUSCHEL (Fig. 24). In order to explain this apparent contradiction, the integration of an extra amino acid into the C-terminal turn of helix I may result in its broadening to the point of the disintegration of the helical conformation and thus shortening of helix I in case of WUS HD and possibly other 66-amino acid WOX homeodomains as well. Furthermore this concludes that structural characteristics others than the length of helix I may not be affected by whether the WOX homeodomain is composed of 66 or 65 amino acids.

A large impact on the loop conformation may be exerted by the identity of the amino acid at position 25. The structural analysis of the WUS HD presented in this study suggests that the conformation of the loop strongly depends on the hydrophobic network orchestrated by Ile25 (section 3.2.3.2). At the corresponding position of WOX13-clade homeodomains only amino acids with short polar side chains are present (Fig. 24), neither of which is likely to interact with the hydrophobic core. Thus the loop conformation of WOX13-clade homeodomains is likely to differ substantially from the respective region in the WUS HD structure. With regard to its conserved function in DNA recognition, differences in loop conformation may also affect the DNA binding properties of the according homeodomain.

Aside of position 25 of the WOX homeodomain, other amino acid residues seem to have been subjected to changes in the course of the evolution. Many correspond to key residues important for homeodomain DNA recognition, which were shown repeatedly to affect *in vitro* and *in vivo* function of homeodomains (vertical lines at the top of Fig. 24, section 1.1.3). Hence differing amino acids at these residues seem very likely to be contributing to a functional divergence between the three clades, which has been indicated in previous publications (section 1.2.3).

Two potential backbone contacting amino acids differ between homeodomains of the WUS-clade and the other two clades: 30 (typ28) and 49 (typ43) (Fig. 24 and Fig. SX). Amino acids at corresponding positions in typical homeodomains, Arg28 and Arg43, were shown to largely account for functional differences *in vivo* between the homeodomains of FTZ and

MSH (Furukubo-Tokunaga, Muller et al. 1992) and affect *in vitro* DNA affinity of PBX1 HD (Lu and Kamps 1996) (see also section 1.1.3.3).

Differing amino acids in the N-terminal arm and at potentially base contacting positions of helix III are likely to result in divergent sequence preferences between homeodomains of the individual clades. Argenines or at least basic amino acids at residues 2, 3 and 5 specify nucleotides of the minor groove by a combination of direct base contact and the selective recognition of narrow minor groove widths (Joshi, Passner et al. 2007; Noyes, Christensen et al. 2008; Dror, Zhou et al. 2014). Nearly all other residues of the N-terminal arm, residues 4, 6, 7 and 8, were also shown to greatly influence homeodomain specificity, most likely by determining the conformation of the N-terminal (Ekker, Jackson et al. 1994; Damante, Pellizzari et al. 1996; Joshi, Passner et al. 2007; Noyes, Christensen et al. 2008). Three residues of the N-terminal arm are invariant in all WOX homeodomains, Arg5 (typ5), Trp6 (typ6) and Pro8 (typ8) (Fig. 24 and Fig. SX). Furthermore docking of the N-terminal arm by means of Trp6 (typ6) in combination with the proline residue at position 8 seems to restrict the flexibility of the N-terminal arm (for more details see section 3.2.4.1). These findings indicate that amino acids of the N-terminal arm ranging from Arg5 to the N-terminus of helix I are likely to adopt a similar conformation in all WOX homeodomains, leaving only the four most N-terminal amino acid residues as potential candidates to convey differential DNA sequence specificity. In particular Pro4 (typ4) of WOX9-clade homeodomains and the basic amino acids at position 3 (typ3) of WOX9- and WOX13-clade homeodomains are likely to affect the sequence preferences.

Differences in sequence specificity determinants of helix III are present at position 60 (typ54) of WOX9-clade homeodomains and at positions 52 (typ46) and 53 (typ47) of WOX13-clade homeodomains (Fig. 24). The effect of the exchange of Ala60 (typ54) for Ser60 (typ54) in WOX9-clade homeodomains on sequence preferences may be only marginal, since neither of their side chains is likely to reach the nucleotides indicated by similar amino acids at corresponding positions in EN HD or CLL HD (Fig. 3A) (Fraenkel, Rould et al. 1998; Miyazono, Zhi et al. 2010). A more severe effect on sequence specificity may be observed for the exchanges of Phe52 (typ46) and Tyr53 (typ47) to Tyr52 (typ46) and Asn53 (typ47) in WOX13-clade homeodomains. Given the diverging properties of the individual amino acids it may be fairly safe to assume that they also display differing sequence preferences. Another but more indirect way for amino acids at the homeodomain positions 52 (typ46) and 53 (typ47) to influence the preferred DNA sequence lies in the modulation of the specificity of Gln56 (typ50).

The corresponding residue of other homeodomains, Gln50, is known for establishing fluctuating contacts with two nucleotides (Billeter, Qian et al. 1993). These interactions seem highly susceptible to the influence of adjacent side chains (Noyes, Christensen et al. 2008) (section 1.1.3.1), which in case of WUS HD are Phe52 (typ46) and Tyr53 (typ47) (Fig. 20). Thus the exchange of these two amino acids for Tyr52 (typ46) and Asn53 (typ47) in WOX13-clade homeodomains may have great influence on the DNA sequence preferences, which could possibly affect the determination of all nucleotides ranging from position 4 to 7 of the DNA binding site.

5 References

- Adams, P. D., P. V. Afonine, et al. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 213-221.
- Ades, S. E. and R. T. Sauer (1994). "Differential DNA-binding specificity of the engrailed homeodomain: the role of residue 50." *Biochemistry* **33**(31): 9187-9194.
- Ades, S. E. and R. T. Sauer (1995). "Specificity of minor-groove and major-groove interactions in a homeodomain-DNA complex." *Biochemistry* **34**(44): 14601-14608.
- Afonine, P. V., R. W. Grosse-Kunstleve, et al. (2012). "Towards automated crystallographic structure refinement with phenix.refine." *Acta Crystallogr D Biol Crystallogr* **68**(Pt 4): 352-367.
- Anderson, W. F., D. H. Ohlendorf, et al. (1981). "Structure of the cro repressor from bacteriophage lambda and its interaction with DNA." *Nature* **290**: 754-758.
- Baird-Titus, J. M., K. Clark-Baldwin, et al. (2006). "The solution structure of the native K50 Bicoid homeodomain bound to the consensus TAATCC DNA-binding site." *J Mol Biol* **356**(5): 1137-1151.
- Banerjee-Basu, S., T. Moreland, et al. (2003). "The Homeodomain Resource: 2003 update." *Nucleic Acids Res* **31**(1): 304-306.
- Bellaoui, M., M. S. Pidkowich, et al. (2001). "The Arabidopsis BELL1 and KNOX TALE homeodomain proteins interact through a domain conserved between plants and animals." *Plant Cell* **13**(11): 2455-2470.
- Berger, M. F., G. Badis, et al. (2008). "Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences." *Cell* **133**(7): 1266-1276.
- Bertolino, E., B. Reimund, et al. (1995). "A Novel Homeobox Protein Which Recognizes a TGT Core and Functionally Interferes with a Retinoid-responsive Motif." *Journal of Biological Chemistry* **270**(52): 31178-31188.
- Billeter, M., Y. Q. Qian, et al. (1993). "Determination of the nuclear magnetic resonance solution structure of an Antennapedia homeodomain-DNA complex." *J Mol Biol* **234**(4): 1084-1093.
- Brennan, R. G. and B. W. Matthews (1989). "The helix-turn-helix DNA binding motif." *Journal of Biological Chemistry* **264**(4): 1903-1906.
- Breuninger, H., E. Rikirsch, et al. (2008). "Differential Expression of *WOX* Genes Mediates Apical-Basal Axis Formation in the *Arabidopsis* Embryo." *Developmental Cell* **14**(6): 867-876.
- Bürglin, T. R. (1994). *A comprehensive classification of homeobox genes*. Oxford ; New York, Oxford University Press.
- Busch, W., A. Miotk, et al. (2010). "Transcriptional control of a plant stem cell niche." *Dev Cell* **18**(5): 849-861.
- Chaney, B. A., K. Clark-Baldwin, et al. (2005). "Solution structure of the K50 class homeodomain PITX2 bound to DNA and implications for mutations that cause Rieger syndrome." *Biochemistry* **44**(20): 7497-7511.
- Chaw, S.-M., C. L. Parkinson, et al. (2000). "Seed plant phylogeny inferred from all three plant genomes: Monophyly of extant gymnosperms and origin of Gnetales from conifers." *Proceedings of the National Academy of Sciences* **97**(8): 4086-4091.

- Chu, S. W., M. B. Noyes, et al. (2012). "Exploring the DNA-recognition potential of homeodomains." *Genome Res* **22**(10): 1889-1898.
- Clarke, N. D. (1995). "Covariation of residues in the homeodomain sequence family." *Protein Sci* **4**(11): 2269-2278.
- Clarke, N. D., C. R. Kissinger, et al. (1994). "Structural studies of the engrailed homeodomain." *Protein Sci* **3**(10): 1779-1787.
- Crooks, G. E., G. Hon, et al. (2004). "WebLogo: a sequence logo generator." *Genome Res* **14**(6): 1188-1190.
- D'Elia, A. V., G. Tell, et al. (2001). "Missense mutations of human homeoboxes: A review." *Hum Mutat* **18**(5): 361-374.
- Damante, G., D. Fabbro, et al. (1994). "Sequence-specific DNA recognition by the thyroid transcription factor-1 homeodomain." *Nucleic Acids Res* **22**(15): 3075-3083.
- Damante, G., L. Pellizzari, et al. (1996). "A molecular code dictates sequence-specific DNA recognition by homeodomains." *Embo Journal* **15**(18): 4992-5000.
- Deveaux, Y., C. Toffano-Nioche, et al. (2008). "Genes of the most conserved WOX clade in plants affect root and flower development in Arabidopsis." *BMC Evol Biol* **8**: 291.
- Dolzblasz, A., J. Nardmann, et al. (2016). "Stem Cell Regulation by *Arabidopsis* WOX Genes." *Molecular Plant* **9**(7): 1028-1039.
- Dror, I., T. Zhou, et al. (2014). "Covariation between homeodomain transcription factors and the shape of their DNA binding sites." *Nucleic Acids Res* **42**(1): 430-441.
- Ekker, S. C., D. G. Jackson, et al. (1994). "The degree of variation in DNA sequence recognition among four Drosophila homeotic proteins." *Embo Journal* **13**(15): 3551-3560.
- Emsley, P., B. Lohkamp, et al. (2010). "Features and development of Coot." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 4): 486-501.
- Fraenkel, E. and C. O. Pabo (1998). "Comparison of X-ray and NMR structures for the Antennapedia homeodomain-DNA complex." *Nat Struct Biol* **5**(8): 692-697.
- Fraenkel, E., M. A. Rould, et al. (1998). "Engrailed homeodomain-DNA complex at 2.2 Å resolution: a detailed view of the interface and comparison with other engrailed structures." *J Mol Biol* **284**(2): 351-361.
- Furukubo-Tokunaga, K., S. Flister, et al. (1993). "Functional specificity of the Antennapedia homeodomain." *Proc Natl Acad Sci U S A* **90**(13): 6360-6364.
- Furukubo-Tokunaga, K., M. Muller, et al. (1992). "In vivo analysis of the helix-turn-helix motif of the fushi tarazu homeo domain of Drosophila melanogaster." *Genes Dev* **6**(6): 1082-1096.
- Gehring, W. J., M. Affolter, et al. (1994). "Homeodomain proteins." *Annu Rev Biochem* **63**: 487-526.
- Gehring, W. J., Y. Q. Qian, et al. (1994). "Homeodomain-DNA recognition." *Cell* **78**(2): 211-223.
- Graf, P., A. Dolzblasz, et al. (2010). "MGOUN1 Encodes an Arabidopsis Type IB DNA Topoisomerase Required in Stem Cell Regulation and to Maintain Developmentally Regulated Gene Silencing." *The Plant Cell* **22**(3): 716-728.
- Grant, R. A., M. A. Rould, et al. (2000). "Exploring the role of glutamine 50 in the homeodomain-DNA interface: crystal structure of engrailed (Gln50 --> ala) complex at 2.0 Å." *Biochemistry* **39**(28): 8187-8192.
- Groß-Hardt, R., M. Lenhard, et al. (2002). "WUSCHEL signaling functions in interregional communication during Arabidopsis ovule development." *Genes & Development* **16**(9): 1129-1138.

- Haecker, A., R. Groß-Hardt, et al. (2004). "Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*." *Development* **131**(3): 657.
- Hall, M. N. and A. D. Johnson (1987). "Homeo domain of the yeast repressor alpha 2 is a sequence-specific DNA-binding domain but is not sufficient for repression." *Science* **237**(4818): 1007-1012.
- Hanes, S. D. and R. Brent (1989). "DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9." *Cell* **57**(7): 1275-1283.
- Hirakawa, Y., Y. Kondo, et al. (2010). "TDIF Peptide Signaling Regulates Vascular Stem Cell Proliferation via the *WOX4* Homeobox Gene in *Arabidopsis*." *The Plant Cell* **22**(8): 2618-2629.
- Hiratsu, K., K. Matsui, et al. (2003). "Dominant repression of target genes by chimeric repressors that include the EAR motif, a repression domain, in *Arabidopsis*." *The Plant Journal* **34**(5): 733-739.
- Ji, J., J. Strable, et al. (2010). "*WOX4* Promotes Procambial Development." *Plant Physiology* **152**(3): 1346-1356.
- Joshi, R., J. M. Passner, et al. (2007). "Functional specificity of a Hox protein mediated by the recognition of minor groove structure." *Cell* **131**(3): 530-543.
- Kabsch, W. (2010). "Xds." *Acta Crystallogr D Biol Crystallogr* **66**(Pt 2): 125-132.
- Kieffer, M., Y. Stern, et al. (2006). "Analysis of the Transcription Factor WUSCHEL and Its Functional Homologue in Antirrhinum Reveals a Potential Mechanism for Their Roles in Meristem Maintenance." *The Plant Cell* **18**(3): 560-573.
- Knosp, W. M., C. Saneyoshi, et al. (2007). "Elucidation, quantitative refinement, and in vivo utilization of the *HOXA13* DNA binding site." *Journal of Biological Chemistry* **282**(9): 6843-6853.
- Kuntal, B. K., P. Aparoy, et al. (2010). "EasyModeller: A graphical interface to MODELLER." *BMC Res Notes* **3**: 226.
- Kyte, J. and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* **157**(1): 105-132.
- Laux, T., K. F. Mayer, et al. (1996). "The WUSCHEL gene is required for shoot and floral meristem integrity in *Arabidopsis*." *Development* **122**(1): 87.
- Leliaert, F., H. Verbruggen, et al. (2011). "Into the deep: New discoveries at the base of the green plant phylogeny." *Bioessays* **33**(9): 683-692.
- Li, T., M. R. Stark, et al. (1995). "Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA." *Science* **270**(5234): 262-269.
- Lin, H., L. Niu, et al. (2013). "Evolutionarily conserved repressive activity of *WOX* proteins mediates leaf blade outgrowth and floral organ development in plants." *Proc Natl Acad Sci U S A* **110**(1): 366-371.
- Lohmann, J. U., R. L. Hong, et al. (2001). "A molecular link between stem cell regulation and floral patterning in *Arabidopsis*." *Cell* **105**(6): 793-803.
- Lu, P., G. B. Rha, et al. (2007). "Structural basis of disease-causing mutations in hepatocyte nuclear factor 1beta." *Biochemistry* **46**(43): 12071-12080.
- Lu, Q. and M. P. Kamps (1996). "Structural determinants within Pbx1 that mediate cooperative DNA binding with pentapeptide-containing Hox proteins: proposal for a model of a Pbx1-Hox-DNA complex." *Mol Cell Biol* **16**(4): 1632-1640.
- Mahony, S., P. E. Auron, et al. (2007). "Inferring protein-DNA dependencies using motif alignments and mutual information." *Bioinformatics* **23**(13): i297-304.

- Mathias, J. R., H. Zhong, et al. (2001). "Altering the DNA-binding specificity of the yeast Matalpha 2 homeodomain protein." Journal of Biological Chemistry **276**(35): 32696-32703.
- Matsumoto, N. and K. Okada (2001). "A homeobox gene, PRESSED FLOWER, regulates lateral axis-dependent development of Arabidopsis flowers." Genes & Development **15**(24): 3355-3364.
- Matthews, B. W., D. H. Ohlendorf, et al. (1982). "Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor." Proceedings of the National Academy of Sciences of the United States of America **79**(5): 1428-1432.
- Mayer, K. F., H. Schoof, et al. (1998). "Role of WUSCHEL in regulating stem cell fate in the Arabidopsis shoot meristem." Cell **95**(6): 805-815.
- McCoy, A. J., R. W. Grosse-Kunstleve, et al. (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.
- McGinnis, W., M. S. Levine, et al. (1984). "A conserved DNA sequence in homoeotic genes of the Drosophila Antennapedia and bithorax complexes." Nature **308**(5958): 428-433.
- McNicholas, S., E. Potterton, et al. (2011). "Presenting your structures: the CCP4mg molecular-graphics software." Acta Crystallogr D Biol Crystallogr **67**(Pt 4): 386-394.
- Miller, S., J. Janin, et al. (1987). "Interior and surface of monomeric proteins." J Mol Biol **196**(3): 641-656.
- Miyazono, K., Y. Zhi, et al. (2010). "Cooperative DNA-binding and sequence-recognition mechanism of aristaless and clawless." Embo Journal **29**(9): 1613-1623.
- Mosca, R., B. Brannetti, et al. (2008). "Alignment of protein structures in the presence of domain motions." BMC Bioinformatics **9**: 352.
- Mosca, R. and T. R. Schneider (2008). "RAPIDO: a web server for the alignment of protein structures in the presence of conformational changes." Nucleic Acids Res **36**(Web Server issue): W42-46.
- Mukherjee, K., L. Brocchieri, et al. (2009). "A comprehensive classification and evolutionary analysis of plant homeobox genes." Molecular Biology and Evolution **26**(12): 2775-2794.
- Nakata, M., N. Matsumoto, et al. (2012). "Roles of the Middle Domain-Specific WUSCHEL-RELATED HOMEBOX Genes in Early Development of Leaves in Arabidopsis." The Plant Cell **24**(2): 519-535.
- Nardmann, J., J. Ji, et al. (2004). "The maize duplicate genes *narrow sheath1* and *narrow sheath2* encode a conserved homeobox gene function in a lateral domain of shoot apical meristems." Development **131**(12): 2827.
- Nardmann, J., P. Reisewitz, et al. (2009). "Discrete Shoot and Root Stem Cell-Promoting WUS/WOX5 Functions Are an Evolutionary Innovation of Angiosperms." Molecular Biology and Evolution **26**(8): 1745-1755.
- Nardmann, J. and W. Werr (2006). "The Shoot Stem Cell Niche in Angiosperms: Expression Patterns of WUS Orthologues in Rice and Maize Imply Major Modifications in the Course of Mono- and Dicot Evolution." Molecular Biology and Evolution **23**(12): 2492-2504.
- Nardmann, J. and W. Werr (2012). "The invention of WUS-like stem cell-promoting functions in plants predates leptosporangiate ferns." Plant Mol Biol **78**(1-2): 123-134.
- Noyes, M. B., R. G. Christensen, et al. (2008). "Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites." Cell **133**(7): 1277-1289.

- Passner, J. M., H. D. Ryoo, et al. (1999). "Structure of a DNA-bound Ultrabithorax-Extradenticle homeodomain complex." *Nature* **397**(6721): 714-719.
- Pellizzari, L., G. Tell, et al. (1997). "Functional interference between contacting amino acids of homeodomains." *FEBS Lett* **407**(3): 320-324.
- Perales, M., K. Rodriguez, et al. (2016). "Threshold-dependent transcriptional discrimination underlies stem cell homeostasis." *Proceedings of the National Academy of Sciences* **113**(41): E6298-E6306.
- Piper, D. E., A. H. Batchelor, et al. (1999). "Structure of a HoxB1-Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation." *Cell* **96**(4): 587-597.
- Qian, Y. Q., M. Billeter, et al. (1989). "The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: Comparison with prokaryotic repressors." *Cell* **59**(3): 573-580.
- Rodriguez, K., M. Perales, et al. (2016). "DNA-dependent homodimerization, sub-cellular partitioning, and protein destabilization control WUSCHEL levels and spatial patterning." *Proceedings of the National Academy of Sciences* **113**(41): E6307-E6315.
- Sakakibara, K., P. Reisewitz, et al. (2014). "*WOX13-like* genes are required for reprogramming of leaf and protoplast cells into stem cells in the moss *Physcomitrella patens*." *Development* **141**(8): 1660.
- Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." *J Mol Biol* **234**(3): 779-815.
- Sarkar, A. K., M. Luijten, et al. (2007). "Conserved factors regulate signalling in Arabidopsis thaliana shoot and root stem cell organizers." *Nature* **446**(7137): 811-814.
- Sato, K., M. D. Simon, et al. (2004). "Dissecting the Engrailed homeodomain-DNA interaction by phage-displayed shotgun scanning." *Chem Biol* **11**(7): 1017-1023.
- Schena, M. and R. W. Davis (1992). "HD-Zip proteins: members of an Arabidopsis homeodomain protein superfamily." *Proc Natl Acad Sci U S A* **89**(9): 3894-3898.
- Schneider, T. D. and R. M. Stephens (1990). "Sequence logos: a new way to display consensus sequences." *Nucleic Acids Res* **18**(20): 6097-6100.
- Schulz, B., F. Banuett, et al. (1990). "The b alleles of U. maydis, whose combinations program pathogenic development, code for polypeptides containing a homeodomain-related motif." *Cell* **60**(2): 295-306.
- Scott, M. P. and A. J. Weiner (1984). "Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax, and fushi tarazu loci of Drosophila." *Proc Natl Acad Sci U S A* **81**(13): 4115-4119.
- Shimizu, R., J. Ji, et al. (2009). "Tissue specificity and evolution of meristematic WOX3 function." *Plant Physiol* **149**(2): 841-850.
- Smith, H. M., I. Boschke, et al. (2002). "Selective interaction of plant homeodomain proteins mediates high DNA-binding affinity." *Proc Natl Acad Sci U S A* **99**(14): 9579-9584.
- Soltis, D. E., C. D. Bell, et al. (2008). "Origin and Early Evolution of Angiosperms." *Annals of the New York Academy of Sciences* **1133**(1): 3-25.
- Stankis, M. M., C. A. Specht, et al. (1992). "The A alpha mating locus of Schizophyllum commune encodes two dissimilar multiallelic homeodomain proteins." *Proc Natl Acad Sci U S A* **89**(15): 7169-7173.
- Stollar, E. J., U. Mayor, et al. (2003). "Crystal structures of engrailed homeodomain mutants: implications for stability and dynamics." *Journal of Biological Chemistry* **278**(44): 43699-43708.

- Stuurman, J., F. Jäggi, et al. (2002). "Shoot meristem maintenance is controlled by a GRAS-gene mediated signal from differentiating cells." Genes & Development **16**(17): 2213-2218.
- Szemenyei, H., M. Hannon, et al. (2008). "TOPLESS Mediates Auxin-Dependent Transcriptional Repression During Arabidopsis Embryogenesis." Science **319**(5868): 1384.
- Tadege, M., H. Lin, et al. (2011). "STENOFOLIA Regulates Blade Outgrowth and Leaf Vascular Patterning in *Medicago truncatula* and *Nicotiana sylvestris*." The Plant Cell **23**(6): 2125-2142.
- Thompson, J. D., D. G. Higgins, et al. (1994). "CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice." Nucleic Acids Res **22**(22): 4673-4680.
- Treisman, J., P. Gonczy, et al. (1989). "A single amino acid can determine the DNA binding specificity of homeodomain proteins." Cell **59**(3): 553-562.
- Tucker-Kellogg, L., M. A. Rould, et al. (1997). "Engrailed (Gln50-->Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity." Structure **5**(8): 1047-1054.
- van der Graaff, E., T. Laux, et al. (2009). "The WUS homeobox-containing (WOX) protein family." Genome Biol **10**(12): 248.
- Vandenbussche, M., A. Horstman, et al. (2009). "Differential recruitment of WOX transcription factors for lateral development and organ fusion in *Petunia* and *Arabidopsis*." Plant Cell **21**(8): 2269-2283.
- Villarejo, M. R. and I. Zabin (1974). "Beta-galactosidase from termination and deletion mutant strains." J Bacteriol **120**(1): 466-474.
- Vollbrecht, E., B. Veit, et al. (1991). "The developmental gene Knotted-1 is a member of a maize homeobox gene family." Nature **350**(6315): 241-243.
- Wickett, N. J., S. Mirarab, et al. (2014). "Phylotranscriptomic analysis of the origin and early diversification of land plants." Proceedings of the National Academy of Sciences of the United States of America **111**(45): E4859-E4868.
- Wilson, D. S., G. Sheng, et al. (1996). "Conservation and diversification in homeodomain-DNA interactions: a comparative genetic analysis." Proc Natl Acad Sci U S A **93**(14): 6886-6891.
- Wolberger, C., A. K. Vershon, et al. (1991). "Crystal structure of a MAT alpha 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions." Cell **67**(3): 517-528.
- Yadav, R. K., M. Perales, et al. (2011). "WUSCHEL protein movement mediates stem cell homeostasis in the *Arabidopsis* shoot apex." Genes Dev **25**(19): 2025-2030.
- Zhang, Y., Y. Jiao, et al. (2017). "Two-Step Functional Innovation of the Stem-Cell Factors WUS/WOX5 during Plant Evolution." Molecular Biology and Evolution **34**(3): 640-653.
- Zhang, Y., C. A. Larsen, et al. (2011). "Structural basis for sequence specific DNA binding and protein dimerization of HOXA13." PLoS One **6**(8): e23069.

6 Supplementary information

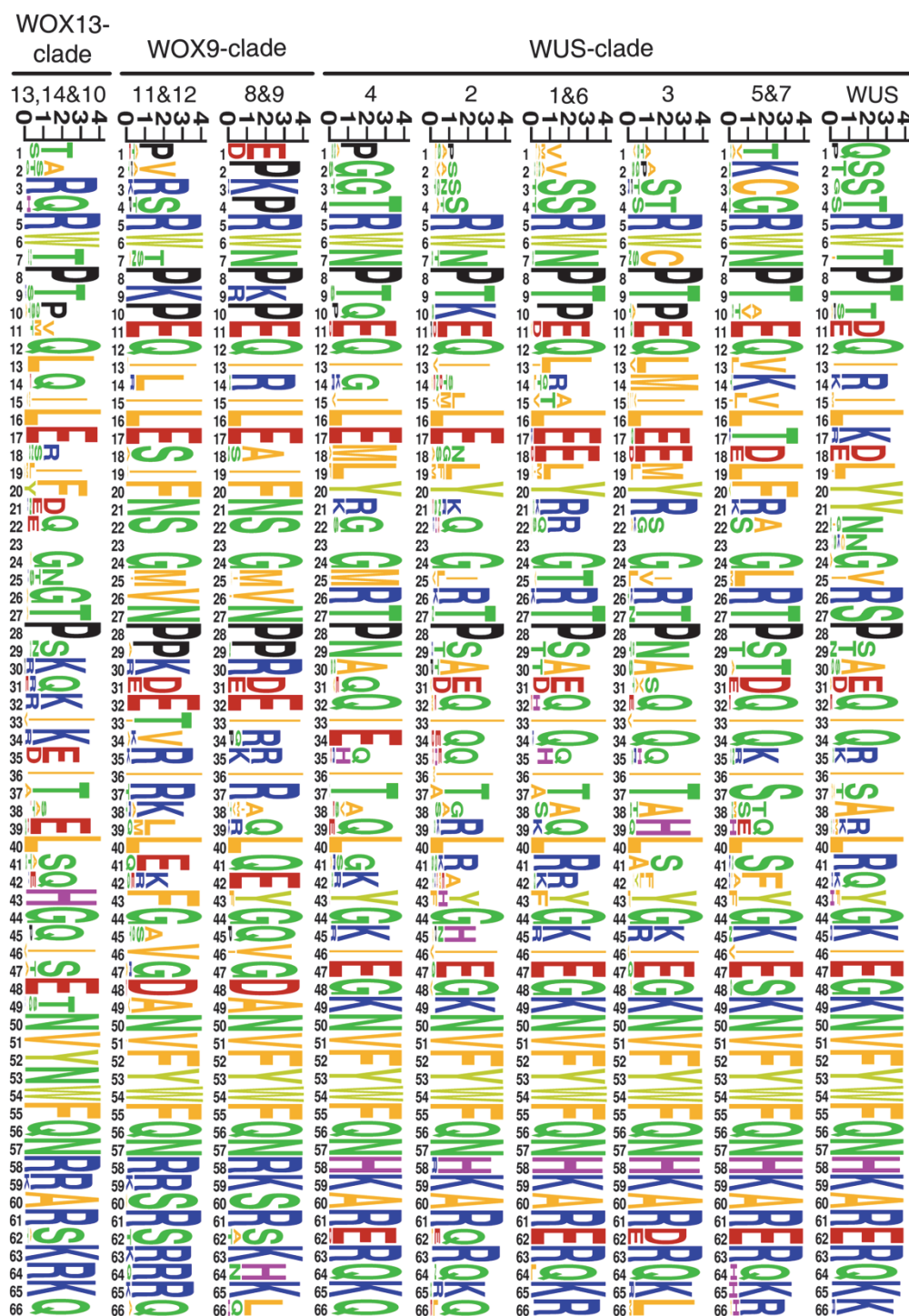


Fig. SX: Amino acid composition of homeodomains of the individual clades as bit scores (y-axis) in function of the homeodomain position (x-axis)

Danksagung

Mein Dank gilt meinem Doktorvater, Prof. Dr. Wolfgang Werr, der mir die Möglichkeit eröffnet hat, in dem für mich damals neuen Feld der Proteinbiochemie und Strukturbioologie zu forschen und wissenschaftliche Erfahrungen zu sammeln.

Des Weiteren möchte ich mich bei meinem Zweitgutachter, Prof. Dr. Ulrich Baumann, und den Mitgliedern seiner Arbeitsgruppe bedanken, für die nette und unkomplizierte Kollaboration, die diese Arbeit überhaupt erst ermöglicht hat.

Auch Dr. Michael Lammers sei an dieser Stelle dafür gedankt, dass ich die ITC Messungen in seinem Labor durchführen durfte und für die Unterstützung, die ich durch ihn und sein Team bei den Experimenten und bei der Auswertung der Daten erfahren habe.

Dr. Viera Kovakov danke ich für die Bereitstellung der Sequenzen der WOX Homöodomänen und die Erlaubnis diese für meine Abbildungen zu verwenden.

Darüber hinaus gebührt mein besonderer Dank Dr. Magdalena Schacherl. Nicht nur hat sie meine Laborarbeit in der Biochemie betreut und die Struktur der WUSCHEL Homöodomäne bestimmt, sondern auch sonst stand sie mir in jeder Hinsicht mit Rat und Tat zur Seite.

Erwähnt seien auch die lieben Kollegen in den diversen Laboren, ohne die meine Zeit in Köln sicherlich nicht halb so schön gewesen wäre.

Last but not least möchte ich mich ganz herzlich bei meinen Eltern, Monika und Peter Koppermann, bedanken, die meinen Werdegang während Studium und Promotion stets befördert haben und mir in jeder Situation ein Rückhalt gewesen sind. Auch meinem Partner und besten Freund, Dr. Martin Lotz, sei für seine stete Unterstützung gedankt.

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Wolfgang Werr betreut worden.